AMERICAN 67 JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXXV

1953

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

PAGE

EMBEDDING THEOREMS FOR ABELIAN GROUPS WITH VALUATIONS.*

By Paul F. Conrad.

Introduction. This study arose from an attempt to simplify the proof and deepen the content of Hahn's Embedding Theorem for ordered abelian groups (Hahn [8]). When doing this, it was discovered that the proper framework for such a discussion is provided by assigning to every element the class of all commensurable elements as its "value," and considering the structure of the groups in terms of the resulting "valuation." An extension of this concept leads to the following general definition of a valuation of an abelian group.

If A is an abelian operator group and Γ is a partially ordered set, then a Γ-*valuation* of A is obtained by assigning to each $a \neq 0$ in A a non-empty trivially ordered subset of Γ, called the set of values of a, subject to the requirement that if none of the values of a and b is greater than γ (greater than or equal to γ), then none of the values of $a \pm b$ and ra is greater than γ (greater than or equal to γ). An abelian operator group with a definite Γ-valuation is called a Γ-*group*. An example of a Γ-group is given by the following direct generalization of the groups introduced by Hahn.

If, to each γ in Γ, there is assigned an abelian operator group $B(\gamma)$ (with respect to a common operator domain R), then the Γ-*sum* V of the $B(\gamma)$ is defined as follows: V is the totality of vectors $b = (\cdots, b_\gamma, \cdots)$, with b_γ in $B(\gamma)$ and $b_\gamma = 0$ for all γ with the exception of a set which satisfies the ascending chain condition. Addition and multiplication (by R) are defined componentwise. The values of b are the maximal γ's with $b_\gamma \neq 0$.

The subgroup C of V is a *c-subgroup* if for every γ in Γ and $b \neq 0$ in $B(\gamma)$, there exists an element c in C with value γ such that $c_\gamma = b$. σ is a Γ-*isomorphism* if σ and σ^{-1} are isomorphisms that preserve values.

In section **3** we prove the following embedding theorem:

If A is a Γ-group over a skew field of operators, then A is Γ-isomorphic to a c-subgroup of a Γ-sum $V(A)$.

In section **4** we prove that a Γ-group over a skew field is a Γ-sum if and

* Received June 22, 1951; revised May 22, 1952.

only if it does not possess any proper c-extensions. We then derive a necessary and sufficient condition for a Γ-isomorphism of A upon B to be extendible to a Γ-isomorphism of A' upon B' where A and B are subgroups of the Γ-sums A' and B' respectively, and the operator domain is a skew field. This enables us to prove that any two embeddings of A into $V(A)$ are essentially the same. In the remainder of Chapter I we use these results to study Γ-automorphisms.

In Chapter II abelian groups without elements of finite order are considered. If such a group is division closed, then it may be considered as an operator group over the field of rational numbers and consequently all of our previous theorems apply. By making use of the well-known theorem that every torsion-free abelian group may be embedded in a division-closed group, we can apply our theory to the general case.

Any partially ordered abelian group may be considered as a Γ-group in a natural way. In this fashion we obtain a simple proof of Hahn's classical embedding theorem for ordered abelian groups and various results that go beyond.

The author wishes to express his appreciation to Reinhold Baer for the suggestions and criticism that he gave during the writing of this paper.

Remark on notation. In this paper *group* will always mean *abelian group,* and the group operation will be denoted by addition. Γ *will always denote a partially ordered set of elements.* That is, a transitive, antisymmetric, reflexive relation $\leq$ is defined between some pairs of elements of Γ. $\alpha, \beta, \gamma, \delta$ will be used to denote the elements of Γ. A subset Φ of Γ is *trivially ordered* if $\alpha \not< \beta$ for α, β in Φ, and Φ is *ordered* (*linearly ordered or simply ordered*) if, for α, β in Φ, either $\alpha \leq \beta$ or $\beta \leq \alpha$.

In the first chapter *all groups will be operator groups with operator domain R.* The operation of R upon a group will be denoted by multiplication from the left. If R is a ring with unit element 1, then $1a = a$, $(rs)a = r(sa)$, and $(r+s)a = ra + sa$ for every a in a group A and every r, s in R. Unless otherwise stated, all subgroups and isomorphisms will be R-subgroups and R-isomorphisms.

Chapter I. Abelian Operator Groups.

1. Γ-groups, Γ-valuations and Γ-sums. There are two equivalent ways of defining a Γ-group. The following definition, equivalent to the one indicated in the introduction, will be most convenient for our applications.

The group A is a Γ-*group* if to every γ in Γ there is assigned a pair of subgroups A_γ, A^γ of A meeting the following requirements:

(a) $A_\gamma \subseteq A^\gamma$ for every γ in Γ.

(b) $\alpha < \beta$ implies $A^\alpha \subseteq A_\beta$.

(c) For every $a \neq 0$ in A there exists at least one γ such that a is in A^γ but not in A_γ.

(d) If a is not in A^γ, then there exists a β such that $\beta > \gamma$ and a is in A^β but not in A_β.

The set of subgroups A_γ, A^γ of A will be called the Γ-*chain* of A and the quotient group A^γ/A_γ will be called the γ-*factor* of A. The set of all γ-factors of A will be denoted by $K(A)$. If no γ-factor of A is 0, and $\alpha \neq \beta$ implies $(A_\alpha, A^\alpha) \neq (A_\beta, A^\beta)$, then A is a *proper* Γ-*group*.

Remark. If (d) is omitted from this definition, then a number of the following theorems are still valid.

γ is a *value* of the element a in the Γ-group A if a is in A^γ but not in A_γ. The set of all values of a will be denoted by Γ^a. Obviously Γ^a is vacuous if and only if $a = 0$. For a Γ-group A we have the following properties.

I. *The element a is in A^γ (A_γ) if and only if none of the values of a is greater than γ (greater than or equal to γ).*

Proof. Let $a \neq 0$ be an element in A and assume that one of the values of a, say β, is greater than γ. It follows from (b) that $A^\gamma \subseteq A_\beta$. Since a is not in A_β, a is not in A^γ. If a is not in A^γ, then by (d) there exists a value β of a such that $\beta > \gamma$. The remainder of the proof is a consequence of the fact that a is in A_γ if and only if a is in A^γ and γ is not a value of a.

II. *If none of the values of a and b is greater than γ (greater than or equal to γ), then none of the values of $a \pm b$ and ra is greater than γ (greater than or equal to γ).*

Proof. This is a consequence of I and the fact that A_γ and A^γ are R-subgroups.

III. *Γ^a is trivially ordered for every $a \neq 0$ in A.*

Proof. Suppose α and β are values of a. If $\alpha < \beta$, then $A^\alpha \subseteq A_\beta$; hence, as a is not in A_β, a is not in A^α either. Therefore α is not a value of a.

Properties I to III show that the values form a valuation as indicated in the introduction. Accordingly we make the following definition.

A Γ-*valuation* of a group A is obtained by assigning to every $a \neq 0$ in A a non-empty trivially ordered subset Γ^a of Γ, called the set of values of a, subject to the requirement that if none of the values of a and b is greater

than γ (greater than or equal to γ), then none of the values of $a \pm b$ and ra is greater than γ (greater than or equal to γ).

THEOREM 1. 1. *A group A with a definite Γ-valuation can be made into a Γ-group by letting*

$A^\gamma =$ *totality of elements, none of whose values is greater than γ.*

$A_\gamma =$ *totality of elements, none of whose values is greater than or equal to γ.*

The valuation defined by this Γ-group is the original one.

Proof. It follows from the definition of a Γ-valuation that A^γ and A_γ are R-subgroups of A for every γ in Γ. It is obvious that $A_\gamma \subseteq A^\gamma$ and that $\alpha < \beta$ implies $A^\alpha \subseteq A_\beta$. If γ is a value of a in A, then a is not in A_γ, but a is in A^γ, since there does not exist any value β of a such that $\beta > \gamma$. Conversely, if a is in A^γ but not in A_γ, then γ is a value of a. Therefore γ is a value of a if and only if a is in A^γ but not in A_γ.

Since the set of values of $a \neq 0$ in A is not empty, there exists at least one γ such that a is in A^γ but not in A_γ. If a is not in A^γ, then there exists a value of a, say β, such that $\beta > \gamma$; hence a is in A^β but not in A_β.

Therefore a Γ-group may be defined directly or in terms of a Γ-valuation, and this fact will be used in the following theory. Unless otherwise stated, *A, B, C, D will always denote Γ-groups,* and *every subgroup S of A will be considered as a Γ-subgroup with Γ-chain*

$$S_\gamma = S \cap A_\gamma \text{ and } S^\gamma = S \cap A^\gamma \text{ for every } \gamma \text{ in } \Gamma.$$

A Γ-subgroup of a proper Γ-group is not necessarily proper.

Construction of Γ-Sum. A subset Φ of Γ satisfies the A. C. C. (ascending chain condition) if *every non-empty subset of Φ contains a maximal element.* We say that *almost every* element γ in a subset Φ of Γ has a certain property (P), if the A. C. C. is satisfied by the set of γ's in Φ that violate (P). Given a set of R-groups $B(\gamma)$ defined for each γ in Γ, let $V = V(\Gamma, B(\gamma))$ be the set of all vectors $b = (\cdot\cdot\cdot, b_\gamma, \cdot\cdot\cdot)$ where b_γ is in $B(\gamma)$ and almost every b_γ is 0. If addition and multiplication (by elements in R) are defined componentwise, then V is a group, since the join of two subsets of Γ which satisfy the A. C. C. also satisfies the A. C. C. *The γ-th component of an element v in V will be denoted by v_γ.* If we define

$V^\gamma =$ totality of elements v in V such that $v_\alpha = 0$ for all $\alpha > \gamma$, and

$V_\gamma =$ totality of elements v in V such that $v_\alpha = 0$ for all $\alpha \geq \gamma$,

then it follows by a straightforward proof that V is a Γ-group, and the values of v are the maximal γ's with $v_\gamma \neq 0$. This Γ-group V will be called *the* Γ*-sum of the* $B(\gamma)$.

We shall denote by V_F the *restricted direct sum* of the $B(\gamma)$, *i. e.*, the totality of all vectors $b = (\cdot\cdot\cdot, b_\gamma, \cdot\cdot\cdot)$ where all but a finite number of the components are zero. Obviously $V_F \subseteq V \subseteq$ complete direct sum of the $B(\gamma)$. If Γ is trivially ordered or if Γ is inversely well ordered, then the Γ-sum V is equal to the complete direct sum of the $B(\gamma)$. If A is a Γ-group, then we can form *the* Γ*-sum of the* γ*-factors* A^γ/A_γ *of* A; and this Γ-sum we shall denote by $V(A)$.

2. Γ-isomorphisms and decompositions. A Γ-*isomorphism* σ of A into B is an isomorphism of A into B with the additional property that γ *is a value of* a *in* A *if and only if* γ *is a value of* $a\sigma$ *in* B. This property is equivalent to

$$A^\gamma\sigma = B^\gamma \cap A\sigma,\ A_\gamma\sigma = B_\gamma \cap A\sigma \text{ for every } \gamma \text{ in } \Gamma.$$

Inverses and products (if defined) of Γ-isomorphisms are Γ-isomorphisms. Isomorphisms will be denoted by π, ρ, σ, τ so that they will be distinguishable from the elements $\alpha, \beta, \gamma, \delta$ of Γ.

Our goal is to prove that if R is a skew field, then A is Γ-isomorphic to a subgroup of $V(A)$. In order to express the fact that $V(A)$ is, in a sense, a minimal containing Γ-sum, we make the following definition. The subgroup B of A is a c*-subgroup of* A (or A is a c*-extension of* B) if $A^\gamma = A_\gamma + B^\gamma$ *for every* γ *in* Γ. A is c*-closed* if there does not exist any proper c-extension of A. Γ-isomorphisms map c-subgroups upon c-subgroups; hence c-closed groups upon c-closed groups. c-subgroups of c-subgroups are c-subgroups. Because of the natural isomorphism

$$A^\gamma/A_\gamma = (A_\gamma + B^\gamma)/A_\gamma \cong B^\gamma/(B^\gamma \cap A_\gamma) = B^\gamma/B_\gamma$$

A and its c-subgroup B have essentially the same γ-factors. If B is a subgroup of A, then for every γ in Γ the *natural isomorphism* $\pi(\gamma)$ of B^γ/B_γ into A^γ/A_γ maps the coset X of B^γ/B_γ upon the coset $X\pi(\gamma) = A_\gamma + X$ of A^γ/A_γ, since $A^\gamma \cap B = B^\gamma$ and $A_\gamma \cap B = B_\gamma$. These $\pi(\gamma)$ induce a Γ-isomorphism π of $V(B)$ into $V(A)$ by the rule that $(v\pi)_\gamma = v_\gamma\pi(\gamma)$ for v in $V(B)$. π will be called the *natural* Γ*-isomorphism of* $V(B)$ *into* $V(A)$.

Clearly the following statements are equivalent:

(a) B *is a* c*-subgroup of* A.

(b) $(B^\gamma/B_\gamma)\pi(\gamma) = A^\gamma/A_\gamma$ *for every* γ *in* Γ.

(c) $V(B)\pi = V(A)$.

To construct the desired Γ-isomorphism of A upon a c-subgroup of $V(A)$ we make use of the following concept.

Definition 2.1. *A set T of subgroups T_γ, defined for every γ in Γ, is a decomposition of A if*

(i) $A_\gamma = A^\gamma \cap T_\gamma$ *for every* γ *in* Γ.

(ii) $A = A^\gamma + T_\gamma$ *for every* γ *in* Γ.

(iii) *Every a in A is in almost every T_γ.*

It is clear that Γ-isomorphisms map decompositions upon decompositions. In section **3** we prove that a Γ-group *over a skew field of operators* possesses a decomposition. The following example shows that this is not true for every Γ-group.

Example 2.2. Let A be the direct sum of C and D where C and D are the additive group of rational numbers, R is the null set, and

$$0 = A_1 \subset A^1 = A_2 = C \subset A^2 = A$$

is the Γ-chain of A. Consider the subgroup B of A that is generated by $(1/2, 1/2)$ and all elements of the form $(0, 1/p)$ where p is an odd prime. If B possesses a decomposition T, then by (i) and (ii) of Definition 2.1, B is the direct sum of B^1 and T_1. But it is easy to prove that $B^1 = A^1 \cap B$ is not a direct summond of B.

LEMMA 2.3. *If B is a subgroup of A, T is a decomposition of A, and S is a decomposition of B such that $S \subseteq T$ (i.e., $S_\gamma \subseteq T_\gamma$ for every γ in Γ), then $S = B \cap T$ (i.e., $S_\gamma = B \cap T_\gamma$ for every γ in Γ). If $B = A$, then $S = T$; hence every decomposition is maximum and minimum.*

Proof.
$$\begin{aligned} T_\gamma \cap B &= T_\gamma \cap (S_\gamma + B^\gamma) = S_\gamma + (T_\gamma \cap B^\gamma) \\ &= S_\gamma + (T_\gamma \cap A^\gamma \cap B^\gamma) = S_\gamma + (A_\gamma \cap B^\gamma) = S_\gamma + B_\gamma = S_\gamma. \end{aligned}$$

If V is a Γ-sum, then the subgroups $N_\gamma =$ *totality of elements v in V such that* $v_\gamma = 0$, form a decomposition N of V. N will be called *the natural decomposition of V.*

THEOREM 2.4. *If T is a decomposition of A, then the mapping $\bar{T}$ of a in A upon $a\bar{T} = (\cdots, (T_\gamma + a) \cap A^\gamma, \cdots)$ is a Γ-isomorphism of A upon the c-subgroup $A\bar{T}$ of $V(A)$. If γ is a value of a in A, then $(a\bar{T})_\gamma = A_\gamma + a$.*

Proof. By a well-known isomorphism law we have

$$A/T_\gamma = (A^\gamma + T_\gamma)/T_\gamma \simeq A^\gamma/(T_\gamma \cap A^\gamma) = A^\gamma/A_\gamma,$$

and we denote by $\tau(\gamma)$ the natural homomorphism of A upon $A^{\gamma}/(T_{\gamma} \cap A^{\gamma})$. For any a in A, $a\bar{T}$ is a well determined element in $V(A)$, since a is in almost every T_{γ}; hence almost every $(a\bar{T})_{\gamma}$ is 0. $(a\bar{T})_{\gamma} = a\tau(\gamma)$ for every γ in Γ. Therefore $\bar{T}$ is a homomorphism, since every $\tau(\gamma)$ is a homomorphism. If $a\bar{T} = 0$, then a is in $\bigcap_{\gamma \varepsilon \Gamma} T_{\gamma}$; but $\bigcap_{\gamma \varepsilon \Gamma} T_{\gamma} = 0$ (since any element $a \neq 0$ in A has at least one value, say α, and a is not in T_{α}). Therefore $\bar{T}$ is an isomorphism of A into $V(A)$.

It follows from the definition of a Γ-group that $A^{\beta} \subseteq \bigcap_{\gamma > \beta} A_{\gamma}$, and since $A_{\gamma} \subseteq T_{\gamma}$ for every γ in Γ, we have $A^{\beta} \subseteq \bigcap_{\gamma > \beta} A_{\gamma} \subseteq \bigcap_{\gamma > \beta} T_{\gamma}$. If a is not in A^{β}, then there exists a value δ of a such that $\delta > \beta$; hence a is not in T_{δ}, so that a is not in $\bigcap_{\gamma > \beta} T_{\gamma}$. Therefore $A^{\beta} = \bigcap_{\gamma > \beta} T_{\gamma}$ and $A_{\beta} = A^{\beta} \cap T_{\beta}$. Hence β is a value of a in A if and only if $(a\bar{T})_{\gamma} = 0$ for all $\gamma > \beta$ and $(a\bar{T})_{\beta} = (T_{\beta} + a) \cap A^{\beta} = A_{\beta} + a \neq A_{\beta}$; thus $\bar{T}$ is a Γ-isomorphism.

For any γ in Γ and any $Z \neq A_{\gamma}$ in A^{γ}/A_{γ}, $Z = A_{\gamma} + a$ where γ is a value of a. $(a\bar{T})_{\gamma} = Z$ and $(a\bar{T})_{\delta} = 0$ for every $\delta > \gamma$; hence $A\bar{T}$ is a c-subgroup of $V(A)$.

Remark. If (d) in the definition of a Γ-group is omitted, then $\bar{T}$ is still an isomorphism and $A^{\gamma}\bar{T} \subseteq V(A)^{\gamma} \cap A\bar{T}$, but in general we cannot claim equality. *$\bar{T}$ will always denote the Γ-isomorphism that is induced by the decomposition T.*

COROLLARY I. *If S and T are any two decompositions of A, then $S = T$ if and only if $\bar{S} = \bar{T}$.*

Proof. The only if is obvious. If $S \neq T$, then we may assume without loss of generality that there exists an a in S_{γ} but not in T_{γ} for some γ in Γ. $(a\bar{S})_{\gamma} = 0$ and $(a\bar{T})_{\gamma} \neq 0$; hence $\bar{S} \neq \bar{T}$.

COROLLARY II. *If B is a subgroup of A, π is the natural Γ-isomorphism of $V(B)$ into $V(A)$, T is a decomposition of A, and S is a decomposition of B such that $S \subseteq T$, then $\bar{T}$ induces $\bar{S}\pi$.*

Proof.
$$\begin{aligned}(b\bar{S})_{\gamma}\pi(\gamma) &= A_{\gamma} + (b\bar{S})_{\gamma} = A_{\gamma} + ((S_{\gamma} + b) \cap B^{\gamma}) \\ &= (A^{\gamma} \cap T_{\gamma}) + ((S_{\gamma} + b) \cap B^{\gamma}) = A^{\gamma} \cap (T_{\gamma} + ((S_{\gamma} + b) \cap B^{\gamma})) \\ &= A^{\gamma} \cap (T_{\gamma} + S_{\gamma} + ((S_{\gamma} + b) \cap B^{\gamma})) = A^{\gamma} \cap (T_{\gamma} + S_{\gamma} + b) \\ &= A^{\gamma} \cap (T_{\gamma} + b) = (b\bar{T})_{\gamma}.\end{aligned}$$

LEMMA 2.5. *If B is a c-subgroup of A and T is a decomposition of A, then $S = B \cap T$ is a decomposition of B.*

Proof. (i) $B^\gamma \cap S_\gamma = B^\gamma \cap (B \cap T_\gamma) = B^\gamma \cap A^\gamma \cap T_\gamma = B^\gamma \cap A_\gamma = B_\gamma$.

$$\text{(ii) } B = B \cap A = B \cap (A^\gamma + T_\gamma) = B \cap (B^\gamma + A_\gamma + T_\gamma)$$
$$= B \cap (B^\gamma + T_\gamma) = B^\gamma + (B \cap T_\gamma) = B^\gamma + S_\gamma.$$

(iii) Every b in B is in almost every S_γ, since an element of B is in S_γ if and only if it is in T_γ.

Remark. If B is merely a subgroup of A, then $B \cap T$ is not necessarily a decomposition of B. For instance, let A and B be the groups in Example 2.2 and let T be the natural decomposition of A.

COROLLARY. *There exists a decomposition of A if and only if A is Γ-isomorphic to a c-subgroup of $V(A)$.*

Proof. The necessity is part of the content of Theorem 2.4. Assume σ is a Γ-isomorphism of A upon a c-subgroup C of $V(A)$. By Lemma 2.5, $N \cap C$ is a decomposition of C where N is the natural decomposition of $V(A)$. σ^{-1} maps $N \cap C$ upon a decomposition of A.

Excursus on ordered Γ. We shall denote by Γ^A, the set of all values of elements in A. If A is a proper Γ-group, then $\Gamma = \Gamma^A$. *Γ^A is ordered if and only if every element $a \neq 0$ in A has at most one value (hence one and only one value).*

Proof. If Γ^A is ordered, then since Γ^a is non-empty and trivially ordered for every $a \neq 0$ in A, Γ^a is a one-element set. Assume that every element in A has at most one value. Consider an element a in A and suppose α is the value of a. If a is not in A_γ, then either γ is a value of a so that $\gamma = \alpha$ or a is not in A^γ, in which case there exists a value of a that is greater than γ; hence $\alpha > \gamma$. Therefore if α is the value of a and $\gamma \not\leq \alpha$, then a is in A_γ. Given α, β such that $A_\alpha \subset A^\alpha$ and $A_\beta \subset A^\beta$, assume $\alpha \not\leq \beta$ and $\beta \not\leq \alpha$. Let a be of value α and b be of value β; then a is in A^α and A_β but not in A_α, and b is in A^β and A_α but not in A_β. Therefore β and α are values of $a + b$, a contradiction.

If Γ^A *is ordered, then the following statements are equivalent:*

(i) *$V(A)$ has no proper c-subgroups.*

(ii) $V(A) = V_F(A)$.

(iii) *Γ^A is well ordered.*

Proof. Let $V = V(A)$. Clearly V_F is a c-subgroup of V; hence (i)

implies (ii). Given $V = V_F$, assume that Γ^A is not well ordered; then there exists an ω^* sequence of elements of Γ^A. But this implies $V \supset V_F$, a contradiction. Therefore (ii) implies (iii).

Suppose that Γ^A is well-ordered, and let B be a proper c-subgroup of V. There exists a γ in Γ^A such that $V^\gamma \supset B^\gamma$ since, otherwise, $V = \bigcup_{\gamma \in \Gamma} V^\gamma = \bigcup_{\gamma \in \Gamma} B^\gamma = B$. Let β be the first element in Γ^A such that $V^\beta \supset B^\beta$. Assume (by way of contradiction) that $V_\beta = B_\beta$. There exists an element a in V^β but not in $B_\beta = V_\beta$. Since B is a c-subgroup of V, there exists an element b in B such that $a \equiv b \bmod V_\beta = B_\beta$. Therefore a is in B, a contradiction. Hence $B_\beta \subset V_\beta$. There exists an element v in V_β but not in B_β. v has one and only one value α, and necessarily $\alpha < \beta$. Therefore $B^\alpha \subset V^\alpha$, contradicting our choice of β; hence (iii) implies (i).

If Γ^A is W. O. and T is a decomposition of A, then $A\bar{T} = V(A)$. For $A\bar{T}$ is a c-subgroup of $V(A)$, and $V(A)$ has no proper c-subgroups.

3. Existence and embedding theorems. In this section we prove that if R is a skew field, then there exists a decomposition of A so that A is Γ-isomorphic to a c-subgroup of $V(A)$. In order to prove this, we make use of the well known complementation theorem: If G is an abelian operator group with respect to a skew field of operators and H is an admissible subgroup of G, then there exists an admissible subgroup J of G such that $G = H \oplus J$.

EXISTENCE THEOREM. *If R is a skew field, B is a subgroup of A, and S is a decomposition of B, then there exists a decomposition T of A such that $S = B \cap T$.*

Proof. There exists a well ordered ascending chain of subgroups $W(x)$ of A such that

(a) $W(0) = B$.

(b) $W(x+1) = W(x) \oplus Ra$ for some $a = a_x$ in A.

(c) If y is a limit ordinal, then $W(y)$ is the join of all $W(x)$ for $x < y$.

(d) $W(z) = A$ for some ordinal z.

We use induction. Suppose that for all $x < y$ we have defined a decomposition $T(x)$ of $W(x)$ such that $T(x) \supseteq T(z)$ for all $z < x$.

Case I. y is a limit ordinal. Define $T(y) = \bigcup_{x<y} T(x)$ (i. e., for any γ in Γ, $T(y)_\gamma$ is the set-theoretical join of all $T(x)_\gamma$ for $x < y$). Clearly $T(y)_\gamma$ is a subgroup of $W(x)$.

(i) $T(y)_\gamma \cap W(y)^\gamma = W(y)_\gamma$.

If c is in $W(y)_\gamma$, then c is in $W(x)$ for some $x < y$; hence c is in $W(x) \cap W(y)_\gamma = W(x)_\gamma \subseteq T(x)_\gamma \subseteq T(y)_\gamma$, and therefore c is in $T(y)_\gamma \cap W(y)^\gamma$. If c is in $T(y)_\gamma \cap W(y)^\gamma$, then c is in $T(x)_\gamma$ for some $x < y$; hence c is in $T(x)_\gamma \cap W(y)^\gamma \subseteq W(x) \cap W(y)^\gamma = W(x)^\gamma$. Thus c is in $T(x)_\gamma \cap W(x)^\gamma = W(x)_\gamma = W(x) \cap W(y)_\gamma$, and therefore c is in $W(y)_\gamma$.

(ii) $T(y)_\gamma + W(y)^\gamma = W(y)$.

Obviously $T(y)_\gamma + W(y)^\gamma \subseteq W(y)$. If c is in $W(y)$, then c is in $W(x)$ for some $x < y$; hence $c = a + b$ where a is in $W(x)^\gamma$ and b is in $T(x)_\gamma$. Since $W(x)^\gamma \subseteq W(y)^\gamma$ and $T(x)_\gamma \subseteq T(y)_\gamma$, c is in $T(y)_\gamma + W(y)^\gamma$.

(iii) Every c in $W(y)$ is in almost every $T(y)_\gamma$.

For if c is in $W(y)$, then c is in $W(x)$ for some $x < y$; hence c is in almost every $T(x)_\gamma$. Since $T(x) \subseteq T(y)$, c is in almost every $T(y)_\gamma$.

Case II. $y = x + 1$. The proof for this case is a consequence of the following lemma.

LEMMA. *If R is a skew field, $C = D \oplus Rc$ for $c \neq 0$ in C and S is a decomposition of D, then a decomposition of C is defined by:*

$$T_\gamma = \begin{cases} C_\gamma + S_\gamma \text{ for } \gamma \text{ in } \Gamma \text{ such that } C = C^\gamma + S_\gamma. \\ S_\gamma + Rc \text{ for } \gamma \text{ in } \Gamma \text{ such that } C \neq C^\gamma + S_\gamma. \end{cases}$$

Proof. If $C = C^\gamma + S_\gamma$ (if and only if $C^\gamma \supset D^\gamma$, since $D^\gamma + S_\gamma = D$ and C "covers" D) then

(i) $C^\gamma \cap T_\gamma = C^\gamma \cap (C_\gamma + S_\gamma) = C_\gamma + (C^\gamma \cap S_\gamma)$

$$= C_\gamma + (C^\gamma \cap D \cap S_\gamma) = C_\gamma + (D^\gamma \cap S_\gamma) = C_\gamma + D_\gamma = C_\gamma$$

and

(ii) $C^\gamma + T_\gamma = C^\gamma + C_\gamma + S_\gamma = C$.

If $C \neq C^\gamma + S_\gamma$ (if and only if $C^\gamma = D^\gamma$ and hence $C_\gamma = D_\gamma$) then

(i) $C^\gamma \cap T_\gamma = C_\gamma$.

$$S_\gamma \subseteq (S_\gamma + Rc) \cap D \subset S_\gamma + Rc.$$

Since $S_\gamma + Rc$ covers S_γ,

$$S_\gamma = (S_\gamma + Rc) \cap D.$$

Therefore

$$C^\gamma \cap T_\gamma = D^\gamma \cap D \cap (S_\gamma + Rc) = D^\gamma \cap S_\gamma = D_\gamma = C_\gamma.$$

(ii) $C^\gamma + T_\gamma = D^\gamma + S_\gamma + Rc = D + Rc = C$.

Assume that c is not in almost every T_γ; then there exist T_{γ_i} $(i = 1, 2, 3, \cdots)$ such that $\gamma_1 < \gamma_2 < \gamma_3 < \cdots$ and c is not in T_{γ_i}. Since c is not in T_{γ_i}, $C = C^{\gamma_i} + S_{\gamma_i}$; hence $c = c_i + s_i$ where c_i is in C^{γ_i} and s_i is in S_{γ_i}. If c_i is in C_{γ_i}, then c is in T_{γ_i}, a contradiction. Therefore γ_i is a value of c_i. $s_1 = c - c_1 = c_i - c_1 + s_i$ is in D. c_1 is in $C^{\gamma_1} \subseteq C_{\gamma_i}$ for $i > 1$. Therefore γ_i is a value of $c_i - c_1$ for $i > 1$; hence s_1 is not in S_i for $i > 1$. Thus S is not a decomposition of D, a contradiction. Therefore c is in almost every T_γ.

Consider any b in C. $b = b' + rc$ where b' is in D and r is in R. Let

$$\Gamma(b) = (\gamma \text{ in } \Gamma \text{ such that } b \text{ is not in } T_\gamma)$$

$$\Gamma(b') = (\gamma \text{ in } \Gamma \text{ such that } b' \text{ is not in } T_\gamma)$$

$$\Gamma(rc) = (\gamma \text{ in } \Gamma \text{ such that } rc \text{ is not in } T_\gamma) = \begin{cases} \text{null set if } r = 0 \\ \Gamma(c) \text{ if } r \neq 0. \end{cases}$$

Therefore $\Gamma(rc)$ satisfies the A. C. C. b' is in T_γ if and only if b' is in S_γ; hence $\Gamma(b')$ satisfies the A. C. C. $\Gamma(b) \subseteq \Gamma(b') \cup \Gamma(rc)$. Therefore $\Gamma(b)$ satisfies the A. C. C.; hence every b in C is in almost every T_γ. This proves the lemma.

The induction is now complete; hence there exists a decomposition T of A such that $T \supseteq S$. By Lemma 2.3, $S = T \cap B$.

By letting $B = 0$ in the existence theorem we have

Corollary I. *If R is a skew field, then any Γ-group possesses a decomposition.*

Embedding Theorem. *If R is a skew field, then there exists a Γ-isomorphism σ of A upon a c-subgroup of $V(A)$ with the additional property that $(a\sigma)_\gamma = A_\gamma + a$ for every a in A and every γ in Γ^a.*

Proof. By Corollary I there exists a decomposition T of A. $\sigma = \bar{T}$ is a Γ-isomorphism of A upon a c-subgroup of $V(A)$ with the desired property.

In order to investigate the class of all Γ-isomorphisms of A upon a c-subgroup of $V(A)$, we first consider the Γ-automorphisms of $V(A)$ that are induced by "automorphisms" of $K(A)$. If, for every γ in Γ, $\sigma(\gamma)$ is an automorphism of A^γ/A_γ, then the set of the $\sigma(\gamma)$ will be called *an automorphism of $K(A)$*. Any Γ-automorphism of A induces an automorphism of $K(A)$. Any automorphism $\sigma(\gamma)$ of $K(A)$ induces a Γ-automorphism σ of $V(A)$ which is defined by the rule

$$(b\sigma)_\gamma = b_\gamma\sigma(\gamma) \text{ for every } b \text{ in } V(A) \text{ and every } \gamma \text{ in } \Gamma.$$

Therefore any Γ-automorphism of A induces a Γ-automorphism of $V(A)$.

A Γ-automorphism of $V(A)$ that is induced by an automorphism of $K(A)$ will be called *special.*

LEMMA 3.1. *If B and C are c-subgroups of the Γ-sum $V = V(\Gamma, B(\gamma))$ and σ is a Γ-isomorphism of B onto C, then the following properties of σ are equivalent:*

(i) *σ is induced by a special Γ-automorphism of V.*

(ii) *For every b in B and every γ, $b_\gamma = 0$ if and only if $(b\sigma)_\gamma = 0$.*

Proof. Clearly (i) implies (ii). Assume (ii) is true. σ induces an automorphism of $K(V)$ and hence an automorphism $\sigma(\gamma)$ of $B(\gamma)$, for every γ in Γ. Let $b \neq 0$ be any element in B and γ any element in Γ such that $b_\gamma \neq 0$. Since B is a c-subgroup of V, there exists an element c in B with $c_\gamma = b_\gamma$, and $c_\alpha = 0$ for every $\alpha > \gamma$. Therefore $(c\sigma)_\gamma = b_\gamma\sigma(\gamma)$. Since $(b-c)_\gamma = 0$, $0 = ((b-c)\sigma)_\gamma = (b\sigma - c\sigma)_\gamma = (b\sigma)_\gamma - (c\sigma)_\gamma$. Therefore $(b\sigma)_\gamma = b_\gamma\sigma(\gamma)$; hence $(b\sigma)_\alpha = b_\alpha\sigma(\alpha)$ for every α in Γ. Thus σ is induced by the $\sigma(\gamma)$.

COROLLARY. *If V is a Γ-sum, σ is a Γ-automorphism of V, and N is the natural decomposition of V, then the following statements are equivalent:*

(i) *σ is special.*

(ii) *σ leaves N invariant.*

(iii) *$b_\gamma = 0$ if and only if $(b\sigma)_\gamma = 0$ for every b in V and every γ.*

This is an immediate consequence of the lemma, where $B = C = V$.

LEMMA 3.2. *If σ is a Γ-isomorphism of A upon a c-subgroup of $V(A)$, then $\sigma = \sigma'\sigma''$ where σ' is induced by a decomposition of A and σ'' is a special Γ-automorphism of $V(A)$.*

Proof. Let N be the natural decomposition of $V(A)$; then by Lemma 2.5, $A\sigma \cap N$ is a decomposition of $A\sigma$. Hence $T = (A\sigma \cap N)\sigma^{-1}$ is a decomposition of A. Let $\sigma' = \bar{T}$; then for any element c in $A\sigma'$ and γ in Γ, the following statements are equivalent: $c_\gamma = 0$; $c\sigma'^{-1}$ is in $T_\gamma = (A\sigma \cap N)_\gamma\sigma^{-1}$; $c\sigma'^{-1}\sigma$ is in $(A\sigma \cap N)_\gamma$; $(c\sigma'^{-1}\sigma)_\gamma = 0$. Therefore by Lemma 3.1, $\sigma'^{-1}\sigma$ is induced by a special Γ-automorphism σ'' of $V(A)$; hence $\sigma = \sigma'\sigma''$.

COROLLARY. *σ is a decomposition-induced Γ-isomorphism if and only if $(a\sigma)_\gamma = A_\gamma + a$ for every a in A and every γ in Γ^a.*

Proof. The necessity of this condition follows from the properties of a

decomposition-induced Γ-isomorphism. Assume the condition is satisfied, let a be any element in A, and suppose γ is in Γ^a; then $A_\gamma + a = (a\sigma)_\gamma = (a\sigma'\sigma'')_\gamma = (A_\gamma + a)\sigma''(\gamma)$. Therefore $\sigma''(\gamma) = 1$; hence $\sigma'' = 1$, so that $\sigma = \sigma'$.

THEOREM 3.3. *If R is a skew field, B is a subgroup of A, π is the natural Γ-isomorphism of $V(B)$ into $V(A)$, and σ is a Γ-isomorphism of B upon a c-subgroup of $V(B)$, then $\sigma\pi$ is induced by a Γ-isomorphism τ of A upon a c-subgroup of $V(A)$.*

Proof. By Lemma 3.2, $\sigma = \bar{S}\rho$ for some decomposition S of B where ρ is a special Γ-automorphism of $V(B)$. By the existence theorem there exists a decomposition T of A such that $S = B \cap T$. By Corollary II of Theorem 2.4 $\bar{T}$ induces $\bar{S}\pi$.

The special Γ-automorphism $\pi^{-1}\rho\pi$ of $V(\Gamma, (A_\gamma + B^\gamma)/A_\gamma)$ can be extended to a special Γ-automorphism $\bar{\rho}$ of $V(A)$. For $A^\gamma/A_\gamma = (A_\gamma + B^\gamma)/A_\gamma \oplus D_\gamma$. Therefore X in A^γ/A_γ has a unique representation $X = Y + Z$ where Y is in $(A_\gamma + B^\gamma)/A_\gamma$ and Z is in D_γ. Define $X\bar{\rho} = Y\pi^{-1}\rho\pi + Z$. Then $\tau = \bar{T}\bar{\rho}$ is a Γ-isomorphism of A upon a c-subgroup of $V(A)$ and τ induces $\bar{S}\pi\pi^{-1}\rho\pi = \bar{S}\rho\pi = \sigma\pi$ on B.

Remark. If $(b\sigma)_\gamma = B_\gamma + b$ for every b in B and every γ in Γ^b, then there exists a τ such that $(a\tau)_\gamma = A_\gamma + a$ for every a in A and every γ in Γ^a. For $\sigma = \bar{S}$ and $\tau = \bar{T}$ will do. Hence in this case we have a restatement of the existence theorem in terms of Γ-isomorphisms.

4. Extension theorem for Γ-isomorphisms.

THEOREM 4.1. *If R is a skew field, then the following properties of a Γ-group A are equivalent.*

(i) *A is c-closed.*

(ii) *$A\bar{T} = V(A)$ for every decomposition T of A.*

(iii) *$A\bar{T} = V(A)$ for at least one decomposition of A.*

(iv) *There exists a Γ-isomorphism of A upon $V(A)$.*

Proof. Assume (i) is true; then $A\bar{T}$ is c-closed for every decomposition T of A. Hence, since $V(A)$ is always a c-extension of $A\bar{T}$, $A\bar{T} = V(A)$. Thus (i) implies (ii). By the existence theorem there exists at least one decomposition of A, hence (ii) implies (iii). Clearly (iii) implies (iv).

To complete the proof, it is sufficient to show that $V(A)$ is c-closed. Let

C be any c-extension of V and N the natural decomposition of V. By the existence theorem there exists a decomposition T of C such that $T \supseteq N$. By Corollary II of Theorem 2.4, $\bar{T}$ induces $\bar{N}\pi$ on V where π is the natural Γ-isomorphism of $V(V)$ upon $V(C)$. Thus $V(C) = V(V)\pi = V\bar{N}\pi = V\bar{T} \subseteq C\bar{T} \subseteq V(C)$; hence $V\bar{T} = C\bar{T}$, and $V = C$.

Remark. It suffices to replace the hypothesis that R is a skew field by the assumption that $V(A)$ is c-closed and A possesses a decomposition.

COROLLARY. *If R is a skew field, then $V(A)$ is c-closed.*

A Γ-isomorphism σ of B upon C induces an isomorphism of $K(B)$ upon $K(C)$ (a set of isomorphisms $\sigma(\gamma)$ of B^γ/B_γ upon C^γ/C_γ for every γ in Γ). The set of $\sigma(\gamma)$ induces a Γ-isomorphism of $V(B)$ upon $V(C)$.

EXTENSION THEOREM FOR Γ-ISOMORPHISMS. *Suppose R is a skew field, B and C are subgroups of the c-closed groups A and D respectively, and σ is a Γ-isomorphism of B upon C. σ can be extended to a Γ-isomorphism of A upon D if and only if the isomorphism of $K(B)$ upon $K(C)$ that is induced by σ can be extended to an isomorphism of $K(A)$ upon $K(D)$.*

Proof.[1] The necessity is trivial. Let S be a decomposition of B; then $T = S\sigma$ is a decomposition of A. Extend S to a decomposition S^* of A and T to a decomposition T^* of D. $\overline{S^*}$ and $\overline{T^*}$ are Γ-isomorphisms onto, by Theorem 4.1. Next let σ induce isomorphisms $\sigma(\gamma)$ of B^γ/B_γ upon C^γ/C_γ which, by hypothesis, are extendable to isomorphisms $\tau(\gamma)$ of A^γ/A_γ upon D^γ/D_γ. These $\tau(\gamma)$'s induce a Γ-isomorphism τ of $V(A)$ onto $V(D)$. $\overline{S^*}\tau\overline{T^*}^{-1}$ is the required Γ-isomorphism of A upon D. To show this we need only show that it induces σ on B. For this it suffices to show $b\overline{S^*}\tau = b\sigma\overline{T^*}$ for every b in B. But $(b\overline{S^*}\tau)_\gamma = (b\bar{S}\tau)_\gamma = (b\bar{S})_\gamma\sigma(\gamma) = (b + s_\gamma + B_\gamma)\sigma(\gamma) = b\sigma + s_\gamma\sigma + C_\gamma$ where s_γ is chosen so that $b + s_\gamma$ is in B^γ and the last step is the definition of $\sigma(\gamma)$. But then $s_\gamma\sigma$ is in T_γ and $b\sigma + s_\gamma\sigma$ is in C^γ, so that $b\sigma + s_\gamma\sigma + C_\gamma = ((b\sigma)\bar{T})_\gamma = ((b\sigma)\overline{T^*})_\gamma$ as desired.

By obvious specializations we obtain the following corollaries (if R is a skew field).

I. *The c-closed groups A and D are Γ-isomorphic if and only if $K(A)$ is isomorphic to $K(D)$.*

II. *If B and C are c-subgroups of the c-closed groups A and D respectively, and σ is a Γ-isomorphism of B upon C, then σ can be extended to a Γ-isomorphism of A upon D.*

[1] The author wishes to thank the referee for suggesting this proof.

III. (a) *Any two c-embeddings of A into $V(A)$ are conjugate.*

(b) *c-closed c-extensions are equivalent.*

Here we say that the subgroups B and C are *conjugate* if there exists a Γ-automorphism σ of A such that $B\sigma = C$. A and D are *equivalent* extensions of B if there exists a Γ-isomorphism σ of A upon D such that $b\sigma = b$ for every b in B. If σ is a Γ-isomorphism of A upon a c-subgroup of $V(A)$, then $A\sigma$ will be called a *c-imbedding* of A into $V(A)$.

5. Groups of Γ-automorphisms. We shall make use the the following notation. $\Delta(A)$ denotes the group of all Γ-automorphisms of A. $\Delta_p(A)$ denotes the group of all Γ-automorphisms of A that induce the identity automorphism on $K(A)$. The Γ-automorphisms belonging to this group will be called *proper.* $\Delta(K)$ denotes the group of all automorphisms of $K(A)$. The mapping of σ in $\Delta(A)$ upon $\overline{\sigma}$ in $\Delta(K)$ (where $\overline{\sigma}$ is the automorphism of $K(A)$ that is induced by σ) is a homomorphism of $\Delta(A)$ into $\Delta(K)$ with kernel $\Delta_p(A)$, hence $\Delta_p(A)$ is normal in $\Delta(A)$. If A is c-closed and R is a skew field, then the homomorphism is onto.

LEMMA 5.1. *If T is a decomposition of A, σ is a Γ-automorphism of A, and τ is the special Γ-automorphism of $V(A)$ that is induced by σ, then $\sigma\bar{T}\tau^{-1}$ is induced by one and only one decomposition of A.*

Proof. $\sigma\bar{T}\tau^{-1}$ is a Γ-isomorphism of A upon a c-subgroup of $V(A)$. If a in A has value γ, then $a\sigma$ has value γ; hence

$$(a\sigma\bar{T}\tau^{-1})_\gamma = (a\sigma\bar{T})_\gamma\sigma^{-1} = ((T_\gamma + a\sigma)\cap A^\gamma)\sigma^{-1} = (A_\gamma + a\sigma)\sigma^{-1} = A_\gamma + a.$$

By the corollary of Lemma 3.2, there exists one (and hence only one) decomposition S of A such that $\bar{S} = \sigma\bar{T}\tau^{-1}$.

Note, if σ is in $\Delta_p(A)$, then τ is the identity; hence $\sigma = \bar{S}\bar{T}^{-1}$.

THEOREM 5.2. *If A possesses a decomposition and $A\bar{T} = V(A)$ for every decomposition T of A, then*

(a) *the set Π of all Γ-isomorphisms of A upon $V(A)$ that are induced by decompositions of A is a "coset" (i. e., if S and T and U are decompositions of A, then $\bar{S}\bar{T}^{-1}\bar{U}$ is in Π).*

(b) *$\Delta_p(A)$ = the totality of all $\bar{S}\bar{T}^{-1}$ for decompositions S, T of A.*

(c) *$\Delta(A)$ = the totality of all $\bar{S}\tau\bar{T}^{-1}$ for decompositions S, T of A and all special Γ-autmorphisms τ of $V(A)$.*

Proof. (a) Let S, T and U be any three decompositions of A; then

$\bar{S}\bar{T}^{-1}\bar{U}$ is a Γ-isomorphism of A upon $V(A)$. If γ is a value of a in A, then $(a\bar{S}\bar{T}^{-1}\bar{U})_\gamma = A_\gamma + a$. Therefore, by the corollary of Lemma 3.2, there exists one and only one decomposition W of A such that $\bar{W} = \bar{S}\bar{T}^{-1}\bar{U}$.

(b) By Lemma 5.1 any proper Γ-automorphism of A is the quotient of two decomposition-induced Γ-isomorphisms of A. Let S, T be any two decompositions of A; then $\bar{S}\bar{T}^{-1}$ is a proper Γ-automorphism of A.

(c) Since A possesses a decomposition, by Lemma 5.1 any Γ-automorphism of A is of the form $\bar{S}\tau\bar{T}^{-1}$ where S and T are decompositions of A and τ is a special Γ-automorphism of $V(A)$. Any such $\bar{S}\tau\bar{T}^{-1}$ is a Γ-automorphism of A.

Remark 1. If R is a skew field, then by Theorem 4.1 and the existence theorem, any c-closed group satisfies the hypotheses of Theorem 5.2.

Remark 2. If the decomposition T of A is kept fixed, then there exists a 1-1 mapping of $\Delta_p(A)$ upon the set Π^* of all decompositions of A. I. e., σ in $\Delta_p(A)$ maps upon S in Π^* such that $\sigma = \bar{S}\bar{T}^{-1}$. If multiplication in Π^* is defined by $U \circ V = W$ where $\bar{W} = \bar{U}\bar{T}^{-1}\bar{V}$, then the mapping is an isomorphism.

In the author's thesis, of which this paper is a part, two applications are made of the preceding theory. First a study is made of conjugate elements and characteristic subgroups of c-closed groups. Then the theory is applied to torsion free abelian groups. The results of the second application are contained in the next chapter.

Chapter II. Abelian Groups Without Elements of Finite Order.

In this chapter we shall attempt to apply the theory of Chapter I to abelian groups without elements of finite order (except 0). Such a group may always be considered as an abelian group with the integers for operators. If in particular such a group A satisfies $A = nA$ for every positive integer n, then A is called division closed (notation *d-closed*); and it is possible to consider in a natural way the field of rational numbers as a field of operators for A. Hence all of the preceding results apply if we restrict our attention to d-closed groups.

In the next section we show that the preceding results can also be applied to groups (and sub-groups) that are not d-closed. In order to do this the definition of valuation has to be slightly amended, since certain properties which we usually deduced from the field properties (in a trivial fashion) now become independent. Accordingly we add to the definition of a Γ-valuation the requirement

(iii) *a and na have the same values for every a and every* $n \neq 0$.

This requirement is easily seen to be equivalent to the property

(e) A_γ *and* A^γ *are relatively d-closed in A for every* γ *in* Γ,

which we add to the definition of a Γ-group. Here we term a subgroup S of A *relatively d-closed* whenever A/S is free of elements of finite order. Hence γ-factors of a Γ-group contain no elements of finite order. Throughout this chapter Γ-group will mean an abelian group without elements of finite order that is also a Γ-group. As before, A, B, C, D will always denote Γ-groups.

1. Generalized existence and embedding theorems. An observant reader will notice that the theory in this section can be extended practically without any change in argument to R-Γ-groups where R is an integral domain and $ra = 0$ implies either $r = 0$ or $a = 0$ (and even more general structures), since such an R can be embedded into one and essentially only one quotient field.

THEOREM 1.1. *For any* Γ*-group B there exists one and essentially only one d-closed* Γ*-group A such that*

(i) *B is a* Γ*-subgroup of A.*

(ii) *If a is in A, then na is in B for some positive integer n.*
(i.e., any other *d-closed* Γ-group that satisfies (i) and (ii) is equivalent to A.) Such an A will be called the *d-closure* of B.

Proof. It is well known that for any abelian group B without elements of finite order there exists one and essentially only one d-closed extension A of B such that (ii) holds. If we define $A_\gamma(A^\gamma)$ to be the d-closure of $B_\gamma(B^\gamma)$ in A, then $(\Gamma, A_\gamma, A^\gamma)$ is a Γ-chain of A and $B_\gamma = B \cap A_\gamma$, $B^\gamma = B \cap A^\gamma$; hence B is a Γ-subgroup of A.

Suppose σ is an isomorphism of A upon a group $C \supseteq B$ such that $b\sigma = b$ for every b in B. Consider any a in A and suppose that α is a value of a; then ther exists a positive integer n such that na is in B. The following statements are equivalent: α is a value of a; α is a value of na; $n(a\sigma) = (na)\sigma = na$ is in B^α but not in B_α; α is a value of $a\sigma$. Therefore σ is a Γ-isomorphism.

LEMMA 1.2. *If every* γ*-factor of B is d-closed and A is the d-closure of B, then A is a c-extension of B.*

Proof. $A^\gamma \supseteq A_\gamma + B^\gamma$. Consider any a in A with value γ. There exists a positive integer n such that na is in B^γ. Since B^γ/B_γ is d-closed, there exists an element c in B^γ such that $nc \equiv na \bmod B_\gamma$. Therefore $n(a - c)$ is in B_γ; hence $a - c$ is in A_γ. Thus a is in $B^\gamma + A_\gamma$.

If A is d-closed, then every γ-factor of A is d-closed. If every γ-factor of A is d-closed, then $V(A)$ is a d-closed group.

If T is a decomposition of A, then the T_γ are relatively d-closed in A. For assume na is in T_γ where a is in A and n is a positive integer; then since $a = a' + t$ where a' is in A^γ and t is in T_γ, $na = na' + nt$. Hence $na' = na - nt$ is in $A^\gamma \cap T_\gamma = A_\gamma$. Therefore a' is in $A_\gamma \subseteq T_\gamma$; hence a is in T_γ. In particular, if A is d-closed, then so are the T_γ.

LEMMA 1. 3. *If A is the d-closure of B and S is a decomposition of B, then the d-closure T of S in A (i. e., $T_\gamma =$ d-closure of S_γ in A) is the one and only decomposition of A such that $S = B \cap T$.*

Proof. (i) $T_\gamma =$ (d-closure of S_γ in A) $\supseteq$ (d-closure of B_γ in A) $= A_\gamma$. Therefore $A_\gamma \subseteq A^\gamma \cap T_\gamma$. Assume there exists an element b of value γ in T_γ. nb is in S_γ for some positive integer n and γ is a value of nb. This contradicts the fact that S is a decomposition of B. Therefore $A_\gamma = A^\gamma \cap T_\gamma$.

(ii) $A \supseteq A^\gamma + T_\gamma$. Let a be any element in A. na is in B for some positive integer n; hence $na = b + s = nb' + ns' = n(b' + s')$ where b is in B^γ, s is in S_γ, b' is in A^γ, and s' is in T_γ. Therefore $a = s' + b'$ is in $A^\gamma + T_\gamma$.

(iii) Every element a in A is in almost every T_γ. For na is in B for some positive integer n; hence na is in almost every S_γ. Thus, as T_γ's are d-closed, a is in almost every T_γ.

Therefore T is a decomposition of A and $T \supseteq S$; hence, by Lemma I: 2. 3,[2] $S = T \cap B$. Consider any decomposition U of A such that $S = U \cap B$. $U_\gamma \supseteq S_\gamma$ and U_γ is d-closed; hence $U_\gamma \supseteq$ d-closure of S_γ in A. Therefore $U \supseteq T$; hence, by Lemma I: 2. 3, $U = T$.

GENERALIZED EXISTENCE THEOREM. *If B is a subgroup of A, every γ-factor of A is d-closed and S is a decomposition of B, then there exists a decomposition T of A such that $S = B \cap T$.*

Proof. Let C be the d-closure of A. $B \subseteq B^* =$ d-closure of B in C. $S \subseteq S^* =$ d-closure of S in B^*. By Lemma II: 1. 3, S^* is a decomposition

[2] Lemma I: 2. 3 denotes Lemma 2. 3 in Chapter I. Similarly, Lemma II: 1. 3 will denote Lemma 1. 3 in Chapter II, etc.

of B^*. By the existence theorem there exists a decomposition U of C such that $S^* = B^* \cap U$. By Lemma II: 1. 2, A is a c-subgroup of C; hence, by Lemma I: 2. 5, $T = A \cap U$ is a decomposition of A.

$$B \cap T = B \cap A \cap U = B \cap U = B \cap B^* \cap U = B \cap S^* = S.$$

Corollary. *If every γ-factor of A is d-closed, then A possesses a decomposition.*

Therefore the hypothesis that A is an R-group and R is a skew field may, wherever it occurs, be replaced by the hypothesis that every γ-factor of A is d-closed. Hence we have the

Generalized Imbedding Theorem. *If every γ-factor of A is d-closed, then there exists a Γ-isomorphism σ of A upon a c-subgroup of $V(A)$ with the additional property that*

$$(a\sigma)_\gamma = A_\gamma + a \text{ for every } a \text{ in } A \text{ and every } \gamma \text{ in } \Gamma^a.$$

Corollary. *If A is any Γ-group and C is its d-closure, then there exists a Γ-isomorphism σ of A upon a subgroup of $V(C)$ such that*

$$(a\sigma)_\gamma = C_\gamma + a \text{ for every } a \text{ in } A \text{ and every } \gamma \text{ in } \Gamma^a.$$

Proof. By the embedding theorem there exists a Γ-isomorphism σ of C upon a c-subgroup of $V(C)$ such that $(c\sigma)_\gamma = C_\gamma + c$ for every c in C and every γ in Γ^c. Therefore $A\sigma$ is a subgroup of $V(C)$. γ is a value of a in A if and only if γ is a value of a in C. Therefore $(a\sigma)_\gamma = C_\gamma + a$ for every a in A and every γ in Γ^a.

Theorem 1. 4. *Every Γ-sum of torsion free abelian groups is c-closed.*

Proof. Let B be a Γ-sum and assume that C is a proper c-extension of B. Let A be the d-closure of C and let N be the natural decomposition of B. $A \supseteq C \supseteq B$. $B\bar{N} = V(B)$. By the existence theorem there exists a decomposition T of A such that $\bar{T}$ induces $\bar{N}\pi$ on B where π is the natural Γ-isomorphism of $V(B)$ into $V(A)$. $B\bar{T} = V(B)\pi = V(\Gamma, (A_\gamma + B^\gamma)/A_\gamma)$, a Γ-sum. Let c be any element in C but not in B. $c\bar{T} = a + b$ where the non-zero components of b are the non-zero components of $c\bar{T}$ that are in $B\bar{T}$ (i. e., of the form $A_\gamma + d$ where d is in B) and the non-zero components of a are the non-zero components of $c\bar{T}$ that are not in $B\bar{T}$. Since $B\bar{T}$ is a Γ-sum, b is in $B\bar{T}$; hence a is in $C\bar{T}$. If $a = 0$, then $c\bar{T}$ is in $B\bar{T}$, a contradiction. Therefore a has at least one value, say γ. Since $B\bar{T}$ is a c-subgroup of $C\bar{T}$, there exists a d in $B\bar{T}$ with $d_\gamma = a_\gamma$. Therefore a_γ is a component in $B\bar{T}$, a contradiction. Thus there does not exist a proper c-extension of B.

THEOREM 1. 5. *If A possesses a decomposition, then the following statement are equivalent*:

(i) *A is c-closed.*

(ii) *$A\bar{T} = V(A)$ for every decomposition T of A.*

(iii) *$A\bar{T} = V(A)$ for at least one decomposition T of A.*

(iv) *There exists a Γ-isomorphism of A upon $V(A)$.*

Proof. This is an immediate consequence of Theorem II: 1. 4 and Theorem I: 4. 1. Obviously the hypothesis that every γ-factor of A is d-closed may be substituted for the given one.

THEOREM 1. 6. *For any A there exists at least one c-closed c-extension.*

Proof. Let A^* be the d-closure of A. By the embedding theorem, $A \subseteq A^* \subseteq V(A^*)$. There exists a maximum c-extension of A in $V(A^*)$, say M. Let M^* be the d-closure of M in $V(A^*)$; then M^* is a c-subgroup of $V(A^*)$.

Let N be any c-extension of M and let N' be the d-closure of N; then $M \subseteq N \subseteq N' \subseteq V(N')$. Let M' be the d-closure of M in $V(N')$; then M' is a c-subgroup of $V(N')$.

By Theorem II: 1. 1 there exists a Γ-isomorphism σ of M' upon M^* such that $m\sigma = m$ for every m in M. By Corollary II of the extension theorem for Γ-isomorphisms, σ can be extended to a Γ-isomorphism of $V(N')$ upon $V(A^*)$. Therefore $N\sigma$ is a c-extension of M in $V(A^*)$; hence, as M is a maximum c-extension of A in $V(A^*)$, $N\sigma = M$. Therefore $N = N\sigma\sigma^{-1} = M\sigma^{-1} = M$.

THEOREM 1. 7. *If B and C are c-closed c-extensions of A and every γ-factor of A is d-closed, then*

(a) *B and C are d-closed groups that are Γ-isomorphic to $V(A)$.*

(b) *B and C are equivalent extensions of A.*

Proof. (a) follows from Theorem II: 1. 5 and the fact that $K(A)$, $K(B)$ and $K(C)$ are essentially the same. (b) follows from Corollary III of the extension theorem for Γ-isomorphisms.

The following example shows that Theorem II: 1. 7 is not true if the hypothesis that every γ-factor of A is d-closed is omitted. Let

$$\Gamma = (-\infty, \cdots, -(i+1), -i, \cdots, -1)$$

be an ordered set of type $1 + \omega^*$. Denote by V the Γ-sum of the groups

$B(\gamma)$ where $B(-\infty) =$ group of integers and $B(-i) =$ group of rationals. Finally let $b(1/2) = (1/2, \cdots, 1/2, \cdots, 1/2)$, and for every odd prime number p let $b(1/p) = (0, \cdots, 1/p, \cdots, 1/p)$. Let us consider the group D that is generated by V_F, $b(1/2)$, and all the $b(1/p)$'s. One can easily verify that

(a) $D^{-\infty}$ is generated by $(1, \cdots, 0, \cdots, 0)$, and

(b) $D^{-\infty}$ is not a direct summond of D.

Therefore D does not possess a decomposition. By Theorem II: 1. 6 there exists a c-closed c-extension A of D. Since D is a c-extension of V_F, A and V are c-closed c-extensions of V_F. Assume (by way of a contradiction) that σ is a Γ-isomorphism of V upon A. Then $N\sigma$ is a decomposition of A; hence $N\sigma \cap D$ is a decomposition of D, a contradiction. Therefore V and A are not Γ-isomorphic.

2. Partially ordered Γ-groups. We begin by stating a number of well known definitions and facts about partially ordered abelian groups. We write p. o. for partially ordered. The group G is a *p. o. group* if G is a p. o. set with respect to a relation $<$, and the following properties are satisfied for every a, b, c *in* G:

(i) $a < b$ implies $a + c < b + c$.

(ii) $na > 0$ for some positive integer n implies $a > 0$.

G is an *o-group* (*simply* or *linearly ordered* group) if for every a and b in G either $a \leq b$ or $b \leq a$.

The set P of positive elements (i. e., elements > 0) of G satisfies

(a) 0 is not in P and P is closed with respect to addition,

(b) P is relatively d-closed in G.

Conversely, assume a set P of elements of a group H satisfies (a) and (b). If we define for every a, b in H: $a < b$ if $b - a$ is in P, then H is a p. o. group and P is the set of positive elements. A subgroup J of G is *convex* if

(a) J is relatively d-closed in G.

(b) If a and c are in J, and $a < b < c$ for b in G, then b is in J.

If J and H are convex subgroups of G, then J *covers* H if $J \supset H$, and for any convex subgroup K of G, $J \supseteq K \supset H$ implies $J = K$. It is clear that the set of all convex subgroups of G is closed under joins of ordered (by inclusion) subsets and under arbitrary intersections.

\6913

Consider the set Γ of all pairs γ of convex subgroups G_γ, G^γ of G such that G^γ covers G_γ. A p. o. of Γ is defined as follows: $\alpha < \beta$ if $G^\alpha \subseteq G_\beta$.

(1) *G is a Γ-group and $(\Gamma, G_\gamma, G^\gamma)$ is a proper Γ-chain of G.*

Proof. $G_\gamma \subset G^\gamma$ for every γ in Γ since G^γ covers G_γ. $\alpha < \beta$ implies $G^\alpha \subseteq G_\beta$ by the definition of p. o. in Γ.

For every $a \neq 0$ in G there exists at least one γ in Γ such that a is in G^γ but not in G_γ. For let G^γ be the intersection of all convex subgroups of G that contain a and let G_γ be a maximal convex subgroup such that $G_\gamma \subseteq G^\gamma$ and a is not in G_γ. The existence of such a maximal subgroup follows from the fact that the join of an ordered subset of convex subgroups is a convex subgroup by the usual arguments. Clearly G^γ covers G_γ and a is in G^γ but not in G_γ.

If a is not in G^γ, then there exists a β in Γ such that $\gamma < \beta$ and a is in G^β but not in G_β. Let G^β be the intersection of all convex subgroups of G that contain G^γ and a, and let G_β be a maximum convex subgroup such that $G^\gamma \subseteq G_\beta \subseteq G^\beta$ and a is not in G_β. Since G^γ and G_γ are convex subgroups, they are relatively d-closed in G.

Remark. By an almost identical proof it can be shown that

(1) For any abelian operator group, the set of all pairs of admissible subgroups J' and J'' such that J'' covers J' forms a proper Γ-chain.

(2) For any abelian lattice-ordered group, the set of all pairs of L-ideals J' and J'' such that J'' covers J' forms a proper Γ-chain.

See Birkhoff [3] for definition and theory of lattice-ordered groups.

Definition 2. 1. *X in G^γ/G_γ is positive if $X \neq G_\gamma$ and nX contains a positive element for some positive integer n.*

(2) *If X in G^γ/G_γ is positive, then mX contains no negative elements for any positive integer m.*

Proof. For some positive integer n, nX contains a positive element, say a. Assume mX contains a negative element b; then nmX contains the negative element nb. Since $nb = h + ma$ where h is in G_γ, $-h > nb - h = ma > 0$. Therefore ma is in G_γ; hence a is in G_γ since G_γ is relatively d-closed. Thus $X = G_\gamma$, a contradiction.

(3) *G^γ/G_γ is p. o. for every γ in Γ.*

Proof. 0 is not positive. X, Y in G^γ/G_γ are positive implies

$X \neq G_\gamma \neq Y$, and there exist positive integers m, n such that mX and nY contain positive elements. Therefore mnX and mnY contain positive elements and by (2) no negative elements; hence $mn(X+Y) \neq G_\gamma$ and contains a positive element. Thus $X+Y$ is positive. If X is in G^γ/G_γ and nX is positive for some positive integer n, then clearly X is positive.

(4) *G^γ/G_γ contains no proper convex subgroups.*

This is a consequence of the fact that G^γ covers G_γ.

(5) *G^γ/G_γ is either trivially ordered or an o-group.*

Proof. Assume G^γ/G_γ possesses a non-trivial p. o. If there exists an $X \neq G_\gamma$ in G^γ/G_γ such that X is neither positive nor negative, then the totality of Y's in G^γ/G_γ such that $nY = mX$ for integers m and $n \neq 0$ is a proper convex subgroup of G^γ/G_γ, which is impossible.

An isomorphism σ of a p. o. group G into a p. o. group H is an *o-isomorphism* if σ and σ^{-1} preserve the given partial orders. We shall make repeated use of the following

THEOREM (Hölder). *If A is an o-group with no proper convex subgroups, then A is o-isomorphic to a subgroup of the additive group of all real numbers.* (For a proof of this see Baer [1].)

Therefore G^γ/G_γ is either trivially ordered or o-isomorphic to a subgroup of the additive group of all real numbers.

(6) *A trivially ordered group H without proper convex subgroups is isomorphic to a subgroup of the additive group of rational numbers.*

Proof. Let h be any non-zero element in H (if no such element exists, then the statement is obviously true). $J = \{g$ in H such that $ng = mh$ for m and $n \neq 0$ integers$\}$ is a non-zero convex subgroup of H; hence $J = H$. The mapping of g in H upon m/n such that $ng = mh$ is an isomorphism of H upon an additive subgroup of the rationals.

(7) *The element a in G is positive if and only if $G_\gamma + a > 0$ for every γ in Γ^a and $a \neq 0$.*

Proof. The necessity follows from the definition of p. o. in G^γ/G_γ. Given $a \neq 0$ in G such that $G_\gamma + a > 0$ for every γ in Γ^a; assume a is not positive. By (2) a is not negative; hence $H = \{b$ in G such that $nb = ma$ for integers m and $n \neq 0\}$ is a convex subgroup that contains a and covers 0. Therefore H is one of the G^γ for γ in Γ^a, but $a \not> 0$, a contradiction. Therefore we have

THEOREM 2.2. *If G is a p. o. group, then*

(i) *The set Γ of all pairs γ of convex subgroups G^γ, G_γ of G such that G^γ covers G_γ form a proper Γ-chain of G when a p. o. in Γ is defined by: $\alpha < \beta$ if $G^\alpha \subseteq G_\beta$.*

(ii) *If a p. o. is defined in G^γ/G_γ by: X in G^γ/G_γ is positive if $X \neq G_\gamma$ and nX contains a positive element for some positive integer n, then G^γ/G_γ is either trivially ordered (hence isomorphic to a subgroup of the additive group of rational numbers) or G^γ/G_γ is o-isomorphic to a subgroup of the additive group of all real numbers.*

(iii) *a in G is positive if and only if $a \neq 0$ and $G_\gamma + a > 0$ for every γ in Γ^a.*

COROLLARY. *The p. o. group G is ordered if and only if*

(a) *Γ (as defined in (i)) is ordered,* and

(b) *G^γ/G_γ is o-isomorphic to a subgroup of the real numbers.*

Proof. If (a) and (b) are satisfied, then $a \neq 0$ in A has one and only one value, say α, and $A_\alpha + a$ is either positive or negative. Therefore $a > 0$ or $a < 0$. If G is ordered, then the set of all convex subgroups of G is ordered by inclusion; hence Γ is ordered. Since every coset of G^γ/G_γ contains a positive element or a negative element, G^γ/G_γ is ordered. Therefore G^γ/G_γ is o-isomorphic to a subgroup of the real numbers.

Definition 2.3. A Γ-group A is a *p. o. Γ-group* if all the γ-factors of A are p. o. If no γ-factor of A contains a proper convex subgroup, then A is a *strictly* p. o. Γ-group or a p. o. Γ-group in the strict sense.

Any Γ-group is a p. o. Γ-group since its γ-factors admit the trivial p. o. By Theorem 2.2 any p. o. group G is a p. o. Γ-group in the strict sense with respect to the set Γ of all pairs of convex subgroups G^γ, G_γ such that G^γ covers G_γ.

THEOREM 2.4. *Suppose that A is a p. o. Γ-group. Define a in A positive if $a \neq 0$ and $A_\gamma + a > 0$ for every γ in Γ^a. Define a in $V(A)$ positive if $a \neq 0$ and $a_\gamma > 0$ for every γ in Γ^a, then*

(a) *A is a p. o. group and A_γ, A^γ are convex subgroups,*

(b) *$V(A)$ is a p. o. group and $V(A)_\gamma$, $V(A)^\gamma$ are convex subgroups,*

(c) *$\bar{T}$ is an o-isomorphism for every decomposition T of A.*

Proof. $a > 0$ and $b > 0$ imply $a + b > 0$ for a, b in A. For let γ be a value of $a + b$ ($a + b$ has a value since otherwise $a = -b$). If neither a nor b is in A^γ, then a has a value $\alpha > \gamma$; hence b is in A^α. If b is in A_α, then α is a value of $a + b$ which is impossible; hence α is a value of b. But this implies $A_\alpha + a + b$ is positive since $A_\alpha + a$ and $A_\alpha + b$ are positive. Hence $a + b$ is not in A^γ which is impossible. Therefore a or b is in A^γ; then a and b are in A^γ. Thus either a or b is not in A_γ; hence $A_\gamma + a + b > 0$.

$na > 0$ for a positive integer n, and a in A implies $a > 0$. For let γ be a value of a; then γ is a value of na. Hence $n(A_\gamma + a) = A_\gamma + na > 0$; hence, since A^γ/A_γ is p. o., $A_\gamma + a > 0$ for every γ in Γ^a. Thus $a > 0$. Therefore, since 0 is not positive, A is a p. o. group.

A^γ is a convex subgroup. For if $c < b < a$ for a, c in A^γ, then $0 < b - c < a - c$ and $0 < a - b$. Therefore $A_\alpha + a - b > 0$ for every α in Γ^{a-b}. Assume b is not in A^γ, then there exists a value β of b such that $\beta > \gamma$. Since $b - c > 0$ and c, a are in A_β, β is in Γ^{a-b} and $0 > A_\beta + c - b = A_\beta + a - b$, a contradiction. A^γ is relatively d-closed in A by the definition of a Γ-group.

By a similar proof it follows that A_γ is convex, hence (a) is true. (b) can be proven by a similar argument.

Let T be any decomposition of A (if A does not possess any decompositions then (c) is obviously true) and consider any element $a \neq 0$ in A. Since γ is a value of a if and only if $(a\bar{T})_\gamma = A_\gamma + a \neq 0$ and $(a\bar{T})_\delta = 0$ for every $\delta > \gamma$, the following statements are equivalent: $a > 0$; $A_\gamma + a > 0$ for every γ in Γ^a; $(a\bar{T})_\gamma > 0$ for every γ in Γ^a; $a\bar{T} > 0$. Therefore (c) is true. Therefore any p. o. Γ-group A is also a p. o. group and any p. o. group can be made into a p. o. Γ-group in the strict sense (where the two Γ's are not necessarily the same). The connecting link is $a > 0$ if and only if $a \neq 0$ and $A_\gamma + a > 0$ for every γ in Γ^a. When discussing p. o. Γ-groups we can, and shall, make use of the fact that they are also p. o. groups.

If B is a subgroup of the p. o. Γ-group A and if we define $B_\gamma + b > 0$ for b in B^γ if $A_\gamma + b > 0$, then B becomes itself a p. o. Γ-group. It follows that if B is a subgroup of the p. o. Γ-group A, then the natural Γ-isomorphism of $V(B)$ into $V(A)$ is an o-isomorphism.

Suppose that A is the d-closure of the p. o. Γ-group B. An element X in A^γ/A_γ has the form $X = A_\gamma + a$ where a is in A^γ and na is in B^γ for some positive integer n. We define X to be positive if $B_\gamma + na$ is positive. Then A is a p. o. Γ-group and B is a p. o. Γ-subgroup of A. As before, there is one and essentially only one such A (with respect to o-isomorphisms) and it will be called the d-closure of B.

If A and B are p. o. Γ-groups, then obviously not every Γ-isomorphism of A into B is an o-isomorphism. Conversely, every o-isomorphism of A into B is not necessarily a Γ-isomorphism. For if all γ-factors of A and B admit only the trivial p. o. then every isomorphism of A into B is an o-isomorphism.

σ is a value-preserving o-isomorphism if σ is a Γ-isomorphism and σ is also an o-isomorphism.

A Γ-isomorphism σ of A into B is a value-preserving o-isomorphism if and only if σ induces an o-isomorphism of $K(A)$ into $K(B)$ (i. e., isomorphism of $K(A)$ into $K(B)$ that preserves the p. o.'s in the γ-factors of A and B). This follows from the fact that $a > 0$ if and only if $a \neq 0$ and $A_\gamma + a > 0$ for every γ in Γ^a. In particular any proper Γ-automorphism of A is a value-preserving o-automorphism. By (c) of Theorem II: 2. 3 every decomposition-induced Γ-isomorphism of a p. o. Γ-group is a value-preserving o-isomorphism. Therefore we have an imbedding theorem for p. o. Γ-groups.

If every γ-factor of the p. o. Γ-group A is d-closed, then there exists a value-preserving o-isomorphism σ of A upon a c-subgroup of $V(A)$ with the additional property that $(a\sigma)_\gamma = A_\gamma + a$ for every a in A and every γ in Γ^a.

By the corollary to the generalized imbedding theorem we have:

For any p. o. Γ-group A there exists a value-preserving o-isomorphism σ of A upon a subgroup of $V(B)$ where B is the d-closure of A. $(a\sigma)_\gamma = B_\gamma + a$ for every a in A and every γ in Γ^a.

If we replace Γ-group by p. o. Γ-group, Γ-isomorphism by value-preserving o-isomorphism, and isomorphism by o-isomorphisms then, with the exception of Theorem I: 3. 3, all the the preceding results and arguments used to prove them are still valid. In Theorem I: 3. 3 we add the hypothesis that A is a p. o. Γ-group in the strict sense.

B is an *a-subgroup* of the Γ-group A (or A is an *a-extension* of B) if $B^\gamma/B_\gamma = 0$ implies $A^\gamma/A_\gamma = 0$. A is a-closed if there does not exist any proper a-extension of A. For o-groups this definition is equivalent to the usual definition of Archimedean extension. An *Archimedean extension* of an o-group G is an o-group $H \supset G$ such that if $h > 0$ is an element in H, then there are $y > 0$ in G and an integer m such that $my > h$ and $mh > y$.

Γ-isomorphisms map a-subgroups upon a-subgroups and hence, a-closed groups upon a-closed groups. An a-extension of an a-extension is an a-exten-

sion. Every c-extension is an a-extension. If A is a-closed, then A is d-closed (since the d-closure of A is an a-extension), and A is c-closed.

In the next four theorems only p. o. Γ-groups in the strict sense will be considered.

For any such group A we define $V_r(A)$ to be the Γ-sum $V(\Gamma, R_\gamma)$ where R_γ is the o-group of reals if the γ-th factor of A is ordered and R_γ is the group of rationals if the γ-th factor of A is trivially ordered. Define v in $V_r(A)$ to be positive if

(a) $v \neq 0$

(b) $R_\gamma = o$-group of reals for every γ in Γ^v

(c) $v_\gamma > 0$ for every γ in Γ^v.

Let $\sigma(\gamma)$ be an o-isomorphism of A^γ/A_γ into the reals if A^γ/A_γ is ordered and an isomorphism of A^γ/A_γ into the rationals if A^γ/A_γ is trivially ordered. The set of $\sigma(\gamma)$ induce a value-preserving o-isomorphism σ of $V(A)$ into $V_r(A)$ which is defined by the rule:

$$(a\sigma)_\gamma = a_\gamma\sigma(\gamma) \text{ for every } a \text{ in } V(A) \text{ and every } \gamma \text{ in } \Gamma.$$

It is obvious that $V(A)\sigma$ is an a-subgroup of $V_r(A)$, and if B is a d-extension of A, then $V_r(A) = V_r(B)$.

THEOREM 2. 5. *If A is a p. o. Γ-group, then*

(a) *there exists a value-preserving o-isomorphism σ of A into $V_r(A)$ and*

(b) *$V_r(A)$ is an a-closed a-extension of $A\sigma$.*

Proof. $A \subseteq A^* = d$-closure of A. By the imbedding theorem, $A \subseteq V(A^*)$. $V(A^*)$ can be embedded into $V_r(A^*) = V_r(A)$ and these imbeddings preserve value and partial order. Clearly A is an a-subgroup of $V_r(A)$.

Let B be any a-extension of $V_r(A) = V$ and let B^* be the d-closure of B. The γ-th factor of B^* is "real" if and only if the γ-th factor of V is "real"; all other γ-factors of V and B^* are "rational." Therefore, the natural Γ-isomorphism π is a value-preserving o-isomorphism of $V(V)$ upon $V(B^*)$. Let N be the natural decomposition of V. By the existence theorem there exists a decomposition T of B^* such that $\bar{T}$ induces $\bar{N}\pi$ on V.

$$V(B^*) = V(V)\pi = V\bar{N}\pi = V\bar{T} \subseteq B^*\bar{T} \subseteq V(B^*);$$

hence $V\bar{T} = B^*\bar{T}$. Therefore $V = B^*$; hence V is a-closed.

By the corollary of Theorem II: 2. 1, for any o-group A the set Γ^A of all pairs γ of convex subgroups A^γ, A_γ such that A^γ covers A_γ is a proper Γ-chain of A. Γ^A is ordered; hence any $a \neq 0$ in A has one and only one value. Moreover each γ-factor is o-isomorphic to a subgroup of the reals. As a corollary of the last theorem we have

HAHN'S THEOREM. *For any o-group A there exists a value-preserving o-isomorphism of A upon an Archimedean-subgroup of* $V(\Gamma^A, R_\gamma)$ *where* $R_\gamma =$ *reals for every* γ *in* Γ^A.

THEOREM 2. 6. *If A is a p. o.* Γ*-group, then the following statements are equivalent:*

(a) *A is a-closed.*

(b) *A is c-closed, every* γ*-factor of A is d-closed, and each non-zero ordered* γ*-factor of A is o-isomorphic to the group of reals.*

Proof. Assume (a) is true; then A is c-closed and A is d-closed. Hence every γ-factor of A is d-closed. By the embedding theorem there exists a value preserving o-isomorphism of A upon $V(A)$. If all ordered γ-factors of A are not o-isomorphic to the group of reals, then there exists a proper a-extension of $V(A)$, a contradiction. Therefore (b) is true.

Assume (b) is true; then, since γ-factors of A are d-closed, there exists a decomposition T of A and since A is c-closed, $A\bar{T} = V(A)$. Clearly $V(A)$ is o-isomorphic to $V_r(A)$ which, by Theorem II: 2. 6, is a-closed. Therefore A is a-closed.

THEOREM 2. 7. *a-closed a-extensions of a p. o.* Γ*-group A are equivalent* (with respect to a value-preserving o-isomorphism).

Proof. Let B and C be any two a-closed a-extensions of A. Let B' be the d-closure of A in B and let C' be the d-closure of A in C. There exists a value-preserving o-isomorphism σ of B' upon C' such that $a\sigma = a$ for every a in A. B' and C' are subgroups of the c-closed groups B and C respectively. The o-isomorphism of $K(B')$ upon $K(C')$ that is induced by σ can be extended to an o-isomorphism of $K(B)$ upon $K(C)$ (since any o-isomorphism of a d-closed subgroup of the reals upon a subgroup of the reals can be extended to an o-automorphism of the reals). Therefore, by the extension theorem for o-isomorphisms, σ can be extended to a value preserving o-isomorphism of B upon C.

The additive group of an ordered field is d-closed, hence possesses a decomposition. This decomposition can be used to prove an embedding theorem for ordered fields, which leads to some interesting structure theory. These results will be in a later paper.

BIBLIOGRAPHY.

[1] R. Baer, "Zur Topologie der Gruppen," *Journal für die Reine und Angewandte Mathematik*, vol. 160 (1929), pp. 208-226.

[2] ———, "The subgroup of the elements of finite order of an abelian group," *Annals of Mathematics*, vol. 37 (1936), pp. 766-781.

[3] G. Birkhoff, "Lattice-ordered groups," *Annals of Mathematics*, vol. 43 (1942), pp. 298-331.

[4] L. W. Cohen and Casper Goffman, "The topology of ordered abelian groups," *Transactions of the American Mathematical Society*, vol. 67 (1949), pp. 310-319.

[5] ———, "On completeness in the sense of Archimedes," *American Journal of Mathematics*, vol. 72 (1950), pp. 747-751.

[6] C. J. Everett and S. Ulam, "On ordered groups," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 208-216.

[7] L. Fuchs, "On the extension of the partial order of groups," *American Journal of Mathematics*, vol. 72 (1950), pp. 191-199.

[8] H. Hahn, "Über die nichtarchimedischen Grössensysteme," *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Vienna*, Section II a, vol. 116 (1907), pp. 601-653.

[9] K. Iwasawa, "On linearly ordered groups," *Journal of Mathematical Society of Japan* (1948), pp. 1-9.

[10] F. W. Levi, "Ordered groups," *Proceedings of the Indian Academy of Sciences*, vol. 16 (1942), pp. 256-263.

[11] ———, "Contributions to the theory of ordered groups," *Proceedings of the Indian Academy of Sciences*, vol. 17 (1943), pp. 199-201.

[12] P. Lorenzen, "Über halbgeordnete Gruppen," *Mathematische Zeitschrift*, vol. 52 (1949).

[13] B. H. Neumann, "On ordered groups," *American Journal of Mathematics*, vol. 71 (1949), pp. 1-18.

METRIC METHODS IN INTEGRAL AND DIFFERENTIAL GEOMETRY.* [1]

By JERRY W. GADDUM.

Introduction. Whatever unity is possessed by the two parts of this paper is seen when it is considered as part of a program of developing geometry by purely metric methods. Of the many writers who have made contributions to metric differential geometry, the reader's attention is called to Menger [6] and [7], Alt [1], Pauc [9], Wald [11], and Blumenthal [3] and [4].

Part I is concerned with arcs in a general metric space. In 1935 W. A. Wilson [12] defined the spread of a mapping f of a metric space X onto a metric space Y. If x, y are elements of X and $f(x)$, $f(y)$ are their images in Y under f, then $\lim f(x)f(y)/xy$ as $x, y \to t$, if this limit exists, is called the spread of f at t. Wilson stated several theorems about the spread of a mapping, but he gave few detailed proofs and no systematic exposition of the properties of the spread.

In Part I we discuss the spread, restricting ourselves to the case where X is the unit interval $[0, 1]$, Y is an arc in a metric space, and f is a homeomorphism of X onto Y. The two principal results of this discussion are 1) a formula for the length of a rectifiable arc in a metric space given explicitly in terms of the defining function, and 2) a sufficient condition that the limit, at a point of an arc, of the ratio between "chord" and arc length be 1.

Part II is devoted to a metric development of the differential geometry of arcs in Euclidean three-dimensional space. The theorems to be proved are, in the main, well known in the case of *analytic* curves, but they have not heretofore been established metrically. The advantages of this approach are twofold. First, we shall be extending the classical results to a wider class of arcs than those which can be represented by analytic functions. Second, by using definition and methods which are purely metric, without any *a priori*

* Received October 23, 1951; revised March 12, 1952.

[1] The preparation of this paper was sponsored (in part) by the Office of the Air Comptroller, USAF.

The writer takes this opportunity to acknowledge his gratitude for Professor L. M. Blumenthal's kind encouragement and advice.

assumptions about differentiability, we may hope to acquire new insight into concepts which are intrinsically geometric.

After necessary preliminary theorems, we prove metrically some properties of bi-regular arcs analogous to properties of analytic arcs. The term *bi-regular* arc, defined in Section **1** of Part II, is an extension of the conventional term *regular arc*. This leads to a metric proof of the first fundamental theorem of curve theory. Then the Frenet-Serret formulas are established, opening the way to use of the tools and results of analysis.

Part I.

1. Definitions and preliminary theorems. Throughout Part I we consider an arc A in a metric space M and a homeomorphism f, of the unit segment $I_1 = [0, 1]$ onto A. If x is a real number in I_1 and p is its image in A under f, we write $p = f(x)$. For $x, y \in I_1$, denote $|x - y|$ by xy and the distance between their images in A by $f(x)f(y)$. If the limit, as x, y approach t, of $f(x)f(y)/xy$ exists, we denote it by $f'(t)$ and call it the spread of f at t.

For $a, b \in I_1$, denote the subarc of A from $f(a)$ to $f(b)$ by $A_f(a, b)$. If $A_f(a, b)$ is rectifiable, its length will be denoted by $s_f(a, b)$, or simply $s(a, b)$ whenever omitting the subscript f will not cause confusion. It is remarked that $s(a, b)$ depends on the mapping f, as does $A(a, b)$, since under a different homeomorphism, $A_g(a, b)$ will in general be a different subarc of A. If A is rectifiable its length is denoted by s.

The following two theorems were formulated by Wilson and are stated here for later use.

THEOREM W_1. *If $f'(t)$ exists at each point of $[a, b] \subset I_1$, the distance quotient $f(x)f(y)/xy$ is bounded on $[a, b]$.*

THEOREM W_2. *If $f'(t)$ exists at each point of $[a, b]$, then $A_f(a, b)$ is rectifiable.*

2. Theorems on the spread. Our first theorem on the spread is a consequence of the use of a double, rather than a single limit, to define the spread.

THEOREM I. 2. 1. *If $f'(t)$ exists at each point of a set $E \subset I_1$, then $f'(t)$ is continuous in E.*

Proof. Given $\epsilon > 0$, let $\delta(t, \epsilon)$ be defined for each $t \in E$ such that

$xt < \delta(t, \epsilon)$, $yt < \delta(t, \epsilon)$ imply that $|f(x)f(y)/xy - f'(t)| < \epsilon$. Suppose $t_0, t_1 \varepsilon E$ with $t_0 t_1 < \delta(t_0, \epsilon)$. Take x, y such that $xt_1 < \delta(t_1, \epsilon)$, $xt_0 < \delta(t_0, \epsilon)$, $yt_1 < \delta(t_1, \epsilon)$, and $yt_0 < \delta(t_0, \epsilon)$. Then

$$\begin{aligned} |f'(t_0) - f'(t_1)| &= |f'(t_0) - f(x)f(y)/xy + f(x)f(y)/xy - f'(t_1)| \\ &\leq |f'(t_0) - f(x)f(y)/xy + f(x)f(y)/xy - f'(t_1)| \\ &< 2\epsilon, \text{ and the theorem is proved.} \end{aligned}$$

The following theorem is a considerable strengthening of a result due to Wilson.

THEOREM I. 2. 2. *If $f'(t)$ exists at each point of a closed set $E \subset I_1$, then the distance quotient $f(x)f(y)/xy$ converges uniformly to $f'(t)$ for $t \varepsilon E$.*

Proof. If we suppose the contrary, then for some $\epsilon > 0$ there is a sequence $\{t_n\}$ of points of E and points x_n, y_n of I_1 such that $x_n t_n < 1/n$, $y_n t_n < 1/n$, and $|f'(t_n) - f(x_n)f(y_n)/x_n y_n| > \epsilon$. Now a subsequence $\{t_{n_i}\}$ exists with a limit point $t \varepsilon E$ and hence $\lim f'(t_{n_i}) = f'(t)$. Also $\lim x_{n_i} = \lim y_{n_i} = t$. But $f'(t) = \lim\limits_{x_{n_i}, y_{n_i} \to t} f(x_{n_i})f(y_{n_i})/x_{n_i} y_{n_i} \neq \lim f'(t_{n_i})$, which contradicts the continuity of f' at t.

THEOREM I. 2. 3. *If $f'(t)$ exists for each $t \varepsilon [a, b]$, then*

$$s(a, b) = \int_a^b f'(t)\, dt.$$

Proof. Given $\epsilon > 0$, let $T_0 = (t_0{}^0, t_1{}^0, \cdots, t_{k_0}{}^0)$, $t_0{}^0 < t_1{}^0 < \cdots < t_{k_0}{}^0$, be an ordered subset of $[a, b]$ such that $\sum\limits_{i=0}^{k_0-1} f(t_i{}^0)f(t_{i+1}{}^0) > s(a, b) - \epsilon$.

Choose $T_1 = (t_0{}^1, t_1{}^1, \cdots, t_{k_1}{}^1) \subset [a, b]$, $t_0{}^1 < t_1{}^1 < \cdots < t_{k_1}{}^1$, such that $T_0 \subset T_1$ and $|\sum\limits_{i=0}^{k_1-1} f'(t_i{}^1) t_i{}^1 t_{i+1}{}^1 - \int_a^b f'(t)\, dt| < \epsilon$. Since T_1 is a refinement of T_0, $\sum\limits_{i=0}^{k_1-1} f(t_i{}^1)f(t_{i+1}{}^1) > s(a, b) - \epsilon$.

By Theorem I. 2. 2, there is a $\gamma > 0$ such that $xy < \gamma$ implies

$$|f'(x) - f(x)f(y)/xy| < \epsilon.$$

Since $f'(t)$ is uniformly continuous on $[a, b]$, there is an $\eta > 0$ such that $t_1 t_2 < \eta$ implies that $|f'(t_1) - f'(t_2)| < \epsilon$. Let $\delta = \min(\gamma, \eta)$.

Choose $T_2 = (t_0{}^2, t_1{}^2, \cdots, t_{k_2}{}^2) \subset [a, b]$, with $t_0{}^2 < t_1{}^2 < \cdots < t_{k_2}{}^2$, such that $T_1 \subset T_2$ and $t_i{}^2 t_{i+1}{}^2 < \delta$, for $i = 0, 1, \cdots, k_2 - 1$. Since $T_1 \subset T_2$

$\sum_{i=0}^{k_2-1} f(t_i^2)f(t_{i+1}^2) > s(a,b) - \epsilon$, and $\left| \sum_{i=0}^{k_2-1} f'(t_i^2)t_i^2 t_{i+1}^2 - \int_a^b f'(t)dt \right| < \epsilon$. Hence

$$| \Sigma f'(t_i^2)\, t_i^2 t_{i+1}^2 - \Sigma f(t_i^2) f(t_{i+1}^2) | < \epsilon(b-a),$$

$$| \Sigma f(t_i^2) f(t_{i+1}^2) - \int_a^b f'(t)\, dt \,| < \epsilon + \epsilon(b-a),$$

and

$$| s(a,b) - \int_a^b f'(t)\, dt \,| < 2\epsilon + \epsilon(b-a).$$

Since ϵ is arbitrary, $s(a,b) = \int_a^b f'(t)\, dt$.

THEOREM I. 2. 4. *If $f'(t)$ exists at each point of $[a,b]$, then for $t \in [a,b]$ with $f'(t) \neq 0$, $\lim_{x,y\to t} f(x)f(y)/s_f(x,y) = 1$.*

Proof.

$$f(x)f(y)/s_f(x,y) - f(x)f(y)/xy \cdot xy/ \int_x^y f'(t)\, dt$$
$$= f(x)f(y)/xy \cdot xy/f'(\xi)xy = f(x)f(y)/xy \cdot 1/f'(\xi),$$

where ξ is between x and y. Hence

$$\lim_{x,y\to t} f(x)f(y)/s_f(x,y)$$
$$= \lim_{x,y\to t} 1/f'(\xi) \cdot \lim_{x,y\to t} f(x)f(y) = 1/f'(t) \cdot f'(t) = 1.$$

Since, as the foregoing discussion shows, $f'(t)$ is very similar to ds/dt of elementary calculus, one is motivated to adopt a weaker definition of the spread, namely, $f'(t) = \lim f(x)f(t)/xt$ as $x \to t$. However, this is not satisfactory. For example, it is a standard theorem in function theory that if the limit of $[\phi(x) - \phi(t)]/x - t$, as $x \to t$, exists, and is continuous at t, it is equal to $\lim [\phi(x) - \phi(y)]/x - y$ as $x, y \to t$. The corresponding theorem is not true, however, for the spread, as can be shown by example. Examples can be found to show that A may be rectifiable although $f'(t)$ fails to exist at a countably infinite set and that A may not be rectifiable although the spread fails to exist at only one point.

Part II.

1. Definitions and some known theorems. We begin with a metric definition of curvature due to Menger [6], and a metric definition of torsion due to Blumenthal [4]. If p_1, p_2, p_3 are three points of a metric space, let

$K(p_1, p_2, p_3)$ denote $[-D(p_1, p_2, p_3)]^{\frac{1}{2}}/p_1p_2 \cdot p_2p_3 \cdot p_1p_3$. If p_1, p_2, p_3 are four points of a metric space, no three of which are linear, let $T(p_1, p_2, p_3, p_4)$ denote the positive square root of

$$T^2(p_1, p_2, p_3, p_4) = \frac{[18 \cdot |D(p_1, p_2, p_3, p_4)|]}{[D(p_1, p_2, p_3) \cdot D(p_1, p_3, p_4) \cdot D(p_1, p_2, p_4) \cdot D(p_2, p_3, p_4)]^{\frac{1}{2}}}.$$

If p is a point of an arc A, we put

$$K(p) = \lim_{p_i \to p} K(p_1, p_2, p_3) \quad \text{and} \quad T(p) = \lim_{p_i \to p} T(p_1, p_2, p_3, p_4),$$

for p_i points of A, and call $K(p)$ and $T(p)$ the metric curvature and torsion, respectively, of A at p.

It is known [10] that in Euclidean space these assume the following forms: $K(p) = \lim_{p_i \to p} [2 \sin p_1; p_2p_3/p_2p_3]$, $T(p) = \lim_{p_i \to p} [3 \sin p_1p_2; p_3p_4/p_3p_4]$, the expression for torsion holding if the curvature exists different from zero at p.

It is further known that if the metric curvature exists at each point of an arc, then the arc is rectifiable. We shall be concerned, unless otherwise stated, with arcs in three-dimensional Euclidean space which possesses at each point metric curvature and torsion, *each different from zero.* Since such an arc is rectifiable, we may suppose it parametrized in terms of arc length, i. e., $x = x(s)$ in vector notation. Such arcs will be called *bi-regular.*

If A is such an arc, it is known that $K(s)$ and $T(s)$ are continuous functions of s and that A possesses at each point a tangent line and an osculating plane [10]. It should be remembered that the tangent line at p is the limiting position of the line $L(p_1, p_2)$ as p_1 and p_2 approach p independently, and that a similar strong limit defines the osculating plane.

Since the existence of the curvature defined above implies the existence of the curvature defined by Haantjes [5] and this in turn implies that the ratio of arc length to chord length, $s(p_1, p_2)/p_1p_2$ approaches unity, i. e., $\lim_{p_i \to p} s(p_1, p_2)/p_1p_2 = 1$, we see that $dx/ds = \alpha$ is a unit vector along the tangent line. In a later section we shall show the existence of a trihedral at each point.

2. Two theorems on Euclidean space. In this section we state two theorems about Euclidean three-dimensional space which will be used in what follows. They are both intuitively evident, and the proofs will be omitted, but they have not been found by the writer in the literature.

We first make some agreements about notation which will be adhered to in what follows. Given points $p_0, p_1, \cdots, p_n, p_{n+1}$ of E_3, the vector $\overrightarrow{p_{i-1} \cdot p_i}$ is denoted by r_i, the vector product of r_i and r_{i+1}, $r_i \times r_{i+1}$ by q_i, the angle $\measuredangle r_i, r_{i+1}$ by θ_i, and the angle $\measuredangle q_i, q_{i+1}$ by ϕ_i. Two polygons P and P' will be said to be similarly oriented if the triple scalar products $[r_i, r_{i+1}, r_{i+2}]$ and $[r_i', r_{i+1}', r_{i+2}']$ have, for each i, the same sign.

THEOREM II. 2. 1. *If two similarly oriented polygons in E_3, P and P', are situated so, that r_1 and r_1' coincide, and q_1 and q_1' coincide, then r_{n+1}, q_n can be carried into r_{n+1}', q_n' by a rotation of P through an angle u, where*

$$u \leq \sum_{i=1}^{n} | \theta_i - \theta_i' | + \sum_{i=1}^{n-1} | \phi_i - \phi_i' |.$$

THEOREM II. 2. 2. *Let π be a plane and θ an angle, not zero, in π. Given $\epsilon > 0$, there is a $\delta > 0$ such that if π' is a plane with $\measuredangle \pi\pi' < \delta$ and θ' the perpendicular projection of θ in π', then $1 - \epsilon < \theta/\theta' < 1 + \epsilon$.*

3. Some theorems on bi-regular arcs. In this section we prove some facts about bi-regular arcs in E_3 which will be used in what follows and which also have some intrinsic value in a systematic development of the metric theory of curves.

If A is a bi-regular arc in E_3 and M is the osculating plane at a point p of A, we are concerned with the perpendicular projection of A on M, or, more particularly, the projection on M of a suitable neighborhood of p. The projection of A on M is clearly a continuous image of A, and we wish to see that there is a neighborhood of p for which the mapping is one-to-one. If we suppose the contrary, there are points p_1 and p_2 arbitrarily close to p which project into the same point. But the plane $\pi(p_1, p_2, p_3)$ is perpendicular to M and this contradicts the fact that M is the osculating plane at p. Hence some neighborhood of p, say A_p, is projected into an arc.

THEOREM II. 3. 1. *The curvature $K(p)$ of a regular arc A is equal to the curvature at p of the perpendicular projection of a suitable neighborhood of p onto the osculating plane at p.*

Proof. Consider a neighborhood of p, say A_p, whose projection is an arc and denote the projection by A_p'.

For a triple q, r, s sufficiently close to p, the plane $\pi(q, r, s)$ is arbitrarily close to M, the osculating plane at p. But $K(q, r, s) = 2 \sin q; r, s/rs$ and $K(q', r', s') = 2 \sin q'; r', s'/r's'$. By Theorem II. 2. 2 the ratio of these

curvatures can be made arbitrarily close to unity by taking points sufficiently close to p, and the theorem follows.

We are now interested in defining a binormal vector. It is to be observed that the existence of the osculating plane is not, of itself, a guarantee that a binormal can be defined so as to be continuous.

We take three points p_1, p_2, p_3 occurring in that order on A. As we have seen, the vector $\overrightarrow{p_1, p_2}/s(p_1, p_2)$ approaches a unit tangent vector α as p_1, p_2 approach p. Letting $r = \overrightarrow{p_1p_2}$, $t = \overrightarrow{p_2p_3}$, if $\lim r \times t/|r| \cdot |t|$ exists, we denote it by γ and call it the binormal at p.

THEOREM II. 3. 2. *If p is a point of a bi-regular arc A, a binormal exists at p.*

We omit the proof, which is straightforward.

Being assured of the existence of a unit binormal γ for bi-regular arcs, we define a unit principal normal $\beta = \gamma \times \alpha$. In a later section we shall prove the Frenet-Serret formulas on the basis of the above definitions.

The notion of similarly oriented polygons was defined in Section 2. We now define a similar concept for arcs. If $\delta > 0$ exists such that for every polygon inscribed in A with distance between consecutive points less than δ, $[r_{i-1}, r_i, r_{i+1}] > 0$ for all i, then the arc is *positively oriented.* If the triple scalar product is negative for all i, the arc is *negatively oriented.*

THEOREM II. 3. 3. *Every bi-regular arc is either positively or negatively oriented.*

Again, we omit the proof.

4. The first fundamental theorem. This section is devoted to a proof of one of the principal results of this paper, namely, *a metric proof of the first fundamental theorem of curve theory for arcs in E_3.* For plane curves, that is, curves whose torsion is identically zero, the theorem was proved by Alt [2] [1] and [2].

We define a special type of polygon called an n-lattice. Let A be an arc with end points a and b, and let $p_0 = a, p_1, \cdots, p_n = b$ be an $(n+1)$-tuple of points of A, occurring in the order given. If $p_0p_1 = p_1p_2 = \cdots = p_{n-1}p_n$, the points are said to form an *n-lattice.* It is known [4] that for every positive integer n, any arc in a metric space contains an n-lattice.

[2] The writer wishes to acknowledge the kindness of Dr. Alt, who made available a copy of his Vienna dissertation.

LEMMA 1. *For a bi-regular arc A and $\epsilon > 0$, there is an N such that, $n > N$, an n-lattice inscribed in A, with side a_n, has the following properties:*

1) for $i = 0, 1, \cdots, n-1$, $1 - \epsilon < s(p_i, p_{i+1})/a_n < 1 + \epsilon$,

2) for $i = 1, 2, \cdots, n-1$, $1 - \epsilon < K(p_i)/[\sin \theta_i / a_n] < 1 + \epsilon$,

3) for $i = 1, 2, \cdots, n-2$, $1 - \epsilon < T(p_i)/[\sin \phi_i / a_n] < 1 + \epsilon$.

Proof. 1) follows since the existence of the metric curvature implies that the limit of the arc-chord ratio is 1.

For n sufficiently large, $1 - \epsilon < K(p_i)/K(p_{i-1}, p_i, p_{i+1}) < 1 + \epsilon$, from the uniform continuity of the curvature. But

$$K(p_{i-1}, p_i, p_{i+1}) = 2 \sin(\pi - \theta_i)/p_{i-1}p_{i+1} = [\sin \theta_i]/a_n \cdot 2a_n/p_{i-1}p_{i+1},$$

and for n sufficiently large, $2a_n/p_{i-1}p_{i+1}$ is arbitrarily close to 1. Hence 2) is true.

The proof of 3) is exactly analogous to that of 2).

LEMMA 2. *For a bi-regular arc A and $\epsilon > 0$, there is an N such that, for $n > N$, an n-lattice inscribed in A has $0 < \theta_i < \epsilon$ and $0 < \phi_i < \epsilon$.*

Proof. Since

$$K(p) = \lim 2 \sin (p_1; p_2, p_3)/p_2p_3 \text{ as } p_i \to p,$$

and

$$T(p) = \lim 3 \sin (p_3, p_4; p_1p_2)/p_1p_2 \text{ as } p_i \to p,$$

the angles θ_i and ϕ_i must have arbitrarily small sines, for n sufficiently large. But since θ_i is the supplement of the vertex angle in an isosceles triangle all of whose angles have sines approaching zero, then $\pi - \theta_i \to \pi$ and $\theta_i \to 0$. Moreover, if the angles ϕ_i approached π, a unique binormal would not exist, in contradiction to Theorem II. 3. 2. Thus the lemma is proved.

We are now ready to prove the fundamental theorem, which may be formulated as follows:

THEOREM II. 4. 1. *If the points of two bi-regular arcs of E_3 can be put into one-to-one correspondence in such a way that the arc lengths between pairs of corresponding points are equal, and the curvature and torsion at corresponding points are equal, then the arcs are congruent.*

If the arcs are A and A^*, denote the correspondence by $p^* = f(p)$, where $p \in A$ and p^* is its correspondent in A^*.

Given $\epsilon > 0$, inscribe n-lattices in A and A^*, denoting the vertices by $p_0, p_1, \cdots, p_n$ and $p_0', p_1', \cdots, p_n'$, respectively. It should be noted that p_i' is not necessarily the same point as $f(p_i)$. However, for n sufficiently large, p_i' is arbitrarily close to $f(p_i)$. Hence we can choose n so large that

$$1 \epsilon < K(p_i)/K(p_i') < 1 + \epsilon \text{ and } 1 - \epsilon < T(p_i)/T(p_i') < 1 + \epsilon.$$

Denoting the sides of the n-lattices by a_n and a_n', respectively, we see that for n sufficiently large,

$$1 - \epsilon < a_n/a_n' < 1 + \epsilon, \text{ since } \lim n \cdot a_n = \lim n \cdot a_n', \text{ where } n \to \infty.$$

We can choose n so large that

$$1 - \epsilon < \sin \phi_i/\sin \phi_i' < 1 + \epsilon \text{ and } 1 - \epsilon < \sin \theta_i/\sin \theta_i' < 1 + \epsilon,$$

since

$$\begin{aligned} \sin \theta_i/\sin \theta_i' &= [\sin \theta_i/a_n]/[\sin \theta_i/a_n'] \cdot a_n/a_n' \\ &= [\sin \theta_i/a_n]/[K(p_i)/(\sin \theta_i/a_n)] \cdot a_n/a_n' \cdot K(p_i')/K(p_i), \end{aligned}$$

and a similar thing is true for $\sin \phi/\sin \phi'$. The desired inequalities follow from Lemma 1 of this section.

Now $\theta_i = \theta_i/\sin \theta_i \cdot \sin \theta_i/a_n \cdot a_n$, and $\phi_i = \phi_i/\sin \phi_i \cdot \sin \phi_i/a_n \cdot a_n$. By Lemma 2 of this section, we can make $\theta_i/\sin \theta_i$ and $\phi_i/\sin \theta_i$ arbitrarily close to 1 by choosing n sufficiently large. Hence we can make n so large that $1 - \epsilon < \theta_i/\theta_i' < 1 + \epsilon$ and $1 - \epsilon < \phi_i/\phi_i' < 1 + \epsilon$.

Furthermore, by Theorem II. 3. 3, we can choose n so large that $[r_{i-1}, r_i, r_{i+1}]$ has the same sign for $i = 1, 2, \cdots, n-1$ and $[r_{i-1}', r_i', r_{i+1}']$ has the same sign for $i = 1, 2, \cdots, n-1$.

In summary, there is an N such that when $n > N$, n-lattices inscribed in A and A^* have the following properties:

a) both lattices are definitely oriented, although perhaps the two are not similarly oriented;

b) $1 - \epsilon < a_n/a_n' < 1 + \epsilon$; c) for each i, $1 - \epsilon < \theta_i/\theta_i' < 1 + \epsilon$ and $1 - \epsilon < \phi_i/\phi_i' < 1 + \epsilon$.

Consider now two such n-lattices, $n > N$, and denote them by L and L' respectively. We wish to show that the distances p_0p_n and $p_0'p_n'$ differ by an amount which approaches zero with ϵ, which will imply that the end points of A and A^* are the same distance apart. But any two points of A and their images in A^* are end points of sub-arcs which satisfy the hypotheses of the theorem, and so the theorem will follow.

If L and L' are not similarly oriented, reflecting L' in the origin gives a lattice congruent to L' and oriented similarly to L. We may suppose, then, that L and L' are similarly oriented, that p_0 and p_0' coincide, that r_1 and r_1' have the same direction, and that q_1 and q_1' have the same direction.

If λ_i denotes the angle between r_i and r_i', then by Theorem II. 2. 1,

$$\lambda_i \leqq \sum_{j=1}^{i=1} | \theta_j - \theta_j' | + \sum_{j=1}^{i=1} | \phi_0 - \phi_0' | \leqq \epsilon \sum_{j=1}^{i=1} \theta_j + \epsilon \sum_{j=1}^{i=1} \phi_j.$$

But $\sum_{j=1}^{n} \theta_j$ is bounded, since $\lim_{n\to\infty} \sum_{j=1}^{n} \theta_j = \lim_{n\to\infty} \sum \theta_j / \sin \theta_j \cdot \sin \theta_j / a_n \cdot a_n$ which is the line integral along A of K. Denoting these line integrals by K_A and T_A respectively, $\lambda_i \leqq \epsilon(K_A + T_A)$.

Now

$$| \sum_{i=1}^{n} r_i - \sum_{i=1}^{n} r_i' | = | \sum_{i=1}^{n} (r_i - r_i') | = | \sum_{i=1}^{n} r_i(1 - \cos \lambda_i) |$$

$$\leq \max_i \, (1 - \cos \lambda_i) \sum_{i=1}^{n} | r_i | \leq \max_i \, (1 - \cos \lambda_i) \cdot s,$$

where s is the length of A. Hence $| \sum r_i - \sum r_i' |$ approaches zero with ϵ; that is, it is arbitrarily small and hence is zero. Thus the theorem is proved.

5. The Frenet-Serret formulas. In this section we establish the Frenet-Serret formulas for a bi-regular arc in E_3. This will answer a number of questions. First, it will establish that if $x = x(s)$ is the vector equation of a bi-regular arc, the functions $x_i(s)$ are three times differentiable. Second, it will furnish the machinery for a proof of the metric version of the second fundamental theorem of curve theory; namely, that given two positive continuous functions $K(s)$, $T(s)$, there is an arc in E_3, uniquely determined except for position and orientation, whose metric curvature and torsion are $K(s)$, $T(s)$, respectively. It will also make possible a metric discussion of arcs in E_3 entirely analogous to the classical discussion.

The proof consists of showing that for bi-regular arcs the metric curvature and torsion are equal to the classical curvature and absolute value of the classical torsion. From this the Frenet-Serret formulas can be derived as in any standard book on differential geometry.

Let A be a bi-regular arc in E_3 and let p be any point of A. We will measure the arc length s from the point p. Let $\theta(s)$ denote the angle between $\alpha(0)$ and $\alpha(s)$, and let $\phi(s)$ denote the angle between $\gamma(0)$ and $\gamma(s)$.

THEOREM II. 5. 1. *The curvature at p is the absolute value of the*

derivative, with respect to arc length, of the angle between the tangent at p and the tangent at a neighboring point. That is, $K(p) = |d\theta/ds|$.

Proof. Choose s so small that the arc from $p(0) = p$ to $p(s)$, which we denote by A_1, has as its projection on the osculating plane at p an arc. In A_1 inscribe an n-lattice with vertices $p_0 = p(0)$; $p_1, \cdots, p_n = p(s)$. Let a be the distance from p_i to p_{i+1} and σ_i the corresponding arc length. Using the same notation as before, $\theta_i = \measuredangle\, r_i, r_{i+1}$, where $r_i = \overrightarrow{p_{i-1}p_i}$.

Define

$$\theta_n' = \sum_{i=1}^{n-1} \theta_i = \sum_{i=1}^{n-1} \alpha_i/\sin\theta_i \cdot \sin\theta_i/a \cdot a/\sigma_i \cdot \sigma_i.$$

Then, defining

$$\theta'(s) = \lim_{s\to\infty} \theta_n', \qquad \theta'(s) = \int_0^s K(s)\,ds.$$

Hence $d\theta'/ds\,\big|_{s=0} = K(p)$.

Now it is not the case that $\theta'(s) = \theta(s)$, but it is true that

$$\lim_{s\to 0} \theta(s)/\theta'(s) = 1.$$

From this the theorem follows, since $\theta(0) = \theta'(0) = 0$.

The above remark is a consequence of the fact that the projection of θ' in the osculating plane at p is equal to the projection of θ. By Theorem II. 2. 2, the ratio of θ to its projection is arbitrarily close to 1 for s sufficiently small.

THEOREM II. 5. 2. *The unit tangent vector* $\alpha(s)$ *has at* p *a derivative whose magnitude is the curvature at* p*, that is,* $K(p) = |d\alpha/ds|$.

Proof. The proof is identical with the classical proof.

THEOREM II. 5. 3. *The vector* $d\alpha/ds$ *lies in the osculating plane.*

Proof. Take $p_0 = p$, $p_1 = p(s_1)$, $p_2 = p(s_2)$. For s_1, s_2 sufficiently small, the unit vectors along p_0p_1 and p_0p_2 are, except perhaps for sense, arbitrarily close to $\alpha(s_1)$ and $\alpha(s_2)$. Hence if they are properly directed, their difference is arbitrarily close to $\Delta\alpha$, and the plane $\pi(p_0, p_1, p_2)$ is arbitrarily close to the vector $\Delta\alpha/\Delta s$. Hence $d\alpha/ds$ is in the osculating plane at p.

Since $d\alpha/ds$ is perpendicular to α and lies in the osculating plane, it lies along the principal normal. By considering the projection in the osculating plane at p of a neighborhood of p whose projection is an arc, and

remembering that the projection lies entirely on one side of the tangent line, we see that α, $d\alpha/ds$, γ form a right-handed system. Since α, β, γ, also form a righthanded system, we have $d\alpha/ds = K\beta$, the first Frenet-Serret formula.

THEOREM II. 5. 4. *The torsion at p is the absolute value of the derivative, with respect to arc length, of the angle between the osculating plane at p and the osculating plane at a neighboring point. That is, $T(p) = | d\phi/ds |$.*

Proof. The proof is the exact analogue of the proof of Theorem II. 5. 1.

THEOREM II. 5. 5. *The unit binormal vector $\gamma(s)$ has at p a derivative whose magnitude is the torsion at p, that is $T(p) = | d\gamma/ds |$.*

Proof. The proof is standard.

Since the metric torsion is non-negative, while the classical torsion is signed, if we want to extend the Frenet-Serret formulas, we must account for the proper sign.

We know that $d\gamma/ds$ lies in the osculating plane, since it is perpendicular to γ, and that its magnitude is $T(p)$. Differentiating the identity $\alpha \cdot \gamma = 0$, where $\alpha \cdot \gamma$ is the scalar product of α and γ, we obtain $d\alpha/ds \cdot \gamma + \alpha \cdot d\gamma/ds = 0$. Since Since $\alpha \cdot d\gamma/ds = 0$, we have $d\gamma/ds = \pm T\beta$. If the curve is positively oriented, we take the $+$ sign, if negatively oriented, we take the $-$ sign.

Two of the Frenet-Serret formulas have now been obtained: $d\alpha/ds = K\beta$, $d\gamma/ds = \pm T\beta$. Differentiating $\beta = \gamma \times \alpha$, we get

$$d\beta/ds = d\gamma/ds \times \alpha + \gamma \times d\alpha/ds = \mp T\gamma - K\alpha.$$

NATIONAL BUREAU OF STANDARDS, LOS ANGELES, CALIFORNIA
AND
UNIVERSITY OF MISSOURI.

BIBLIOGRAPHY.

[1] F. Alt, *Über eine metrische Definition der Krümmung einer Kurve*, Vienna Dissertation (1931).

[2] ———, *Zur Theorie der Krümmung*, Ergebnisse eines mathematischen Kolloquiums (Wien), Heft 3 (1930), p. 5.

[3] L. M. Blumenthal, *Distance Geometries*, University of Missouri Studies, vol. 13 (1938).

[4] ———, *Distance Geometry* (mimeographed, University of Missouri, 1950).

[5] J. Haantjes, "Distance geometry. Curvature in abstract metric spaces," *Academie van Wetenschappen* (*Amsterdam*) *Proceedings*, vol. 50 (1947), pp. 496-508.

[6] K. Menger, *Untersuchungen über allgemeine Metrik.* Vierte Untersuchung. Zur Metrik der Kurven, *Mathematische Annalen*, vol. 103, pp. 466-501.

[7] ———, "La géométrie des distances et ses relations avec les autres branches des mathématiques," *L'Enseignment Mathématique*, No. 5, 6 (1936), pp. 348-372.

[8] C. Pauc, "Courbure dans les espaces métriques," *Rendiconti, Academia Nazionale dei Lincei*, vol. 24 (1936), pp. 109-115.

[9] ———, "Les Méthodes Directes en Géométrie Différentielle," *Ésposés D'Analyse Générale*, XIII (1941).

[10] J. W. Sawyer, *A Study of Metric Torsion*, University of Missouri Dissertation (1951).

[11] A. Wald, *Begründung einer koordinatenlosen Differentialgeogetrie der Flächen*, Ergebnisse eines mathematischen Kolloquiums (Wien), Heft 7 (1936), pp. 24-26.

[12] W. A. Wilson, "On certain types of continuous transformation of metric spaces," *American Journal of Mathematics*, vol. 57 (1935), pp. 62-68.

AN APPROXIMATION TO TRANSONIC FLOW OF A POLYTROPIC GAS.*

By John A. Tierney.

Introduction. The equation of state of a gas whose flow is governed by Frankl's equation [2] $\psi_{\sigma\sigma} + K(\sigma)\psi_{\zeta\theta} = 0$ is determined by the choice of the function $K(\sigma)$. With the proper choice of units, for a polytropic gas $K(\sigma) = \sigma - b\sigma^2 + \cdots$, $b \doteq 2^{-1}(2\gamma + 5)(\gamma + 1)^{-\frac{2}{3}}$ where γ is the adiabatic exponent.

By retaining only the first term in this expansion we obtain $K(\sigma) = \sigma$, and Frankl's equation becomes Tricomi's equation, used by various authors [1, 5] in the study of transonic flows. The gas determined by taking $K(\sigma) = \sigma$ is called the Tricomi gas [7], or a T_1-gas.

In Section 2 we show that if two functions $K(\sigma)$ have the same expansions in positive integral powers of σ up to and including the term in σ^n, the graphs in the (ρ, p)-plane of the corresponding equations of state have contact of order at least $n + 1$ at the sonic point (ρ_*, p_*). Thus for example if we approximate $K(\sigma)$ above for a polytropic gas by retaining only the first two terms in its expansion, the equation of state corresponding to $K(\sigma) = \sigma - b\sigma^2$ can be made to have contact of the third order at the sonic point (ρ_*, p_*). A gas with an equation of state corresponding to $K(\sigma) = \sigma - b\sigma^2$ is called a T_2-gas and in Section 4 we study its equation of state. In the concluding section we study a particular example of transonic flow for a T_2-gas in the physical plane.

1. Preliminary considerations. Employing the pressure p and the stream function ψ as independent variables Martin [6] has shown that, given a Bernoulli function $q = q(p, \psi)$ and a direction function $\theta = \theta(p, \psi)$ which jointly satisfy

$$q[(q_{pp} - q\theta_p^2)/\theta_\psi]_\psi + (q^2\theta_p)_p = 0, \tag{1}$$

a flow is presented in the physical plane $z = x + iy$ by

$$z = \int e^{i\theta}\{[(q\theta_p^2 - q_{pp})/\theta_\psi]dp + [q\theta_p - iq_p]d\psi\}. \tag{2}$$

* Received October 15, 1951.

The density ρ and Mach number M are given by

(3) $$\rho = -(qq_p)^{-1}, \qquad M^2 = 1 + qq_p^{-2}q_{pp}.$$

Irrotational flows are characterized by a Bernoulli function of the form $q = q(p)$ and we see from (3) that q is a decreasing function of p for $\rho > 0$, $q > 0$. Furthermore, the flow is subsonic, sonic, or supersonic according as $q_{pp} \lesseqqgtr 0$. The *sonic speed* is given by $q_* = q(p_*)$, where p_*, the *sonic pressure,* is defined by $q_{pp}(p_*) = 0$.

If we introduce a new variable (cf. [2])

(4) $$\sigma = \int_{p_*}^{p} q^{-2}dp = \sigma(p)$$

in place of p, and set

(5) $$K(\sigma) = -q^3 q_{pp},$$

the variable p being eliminated from the second member with the aid of (4), equation (1) is replaced by

(6) $$\theta_{\sigma\sigma} = [(K + \theta_\sigma^2)/\theta_\psi]_\psi,$$

and (2) by

(7) $$z = \int \chi e^{i\theta}\{[(K + \theta_\sigma^2)/\theta_\psi]d\sigma + [\theta_\sigma + i\chi^{-1}\chi_\sigma]d\psi\},$$

where the function

(8) $$\chi = \chi(\sigma) = q^{-1}$$

is obtained by eliminating p from $q = q(p)$ again with the help of (4).

$K(\sigma)$ as defined by (5) and (4) is identical with $K(\sigma)$ in Frankl's equation $\psi_{\sigma\sigma} + K(\sigma)\psi_{\theta\theta} = 0$, equivalent to (6). It follows from (4), (5), and (8), that

(9) $$\chi_{\sigma\sigma} - K(\sigma)\chi = 0.$$

A solution $\theta = \theta(\sigma, \psi)$ of (6) when inserted in (7) yields a mapping $z = z(\sigma, \psi)$ of the (σ, ψ)-plane upon the physical plane, which carries the straight lines $\sigma =$ const. into the isovels (isobars) and the straight lines $\psi =$ const. into the streamlines.

2. The equation of state. Order of contact. Alternatively, given $K(\sigma)$, corresponding to a solution $\chi(\sigma)$ of (9) we find from (8), (3), and (4) that

(10) $$q = \chi^{-1} = q(\sigma),$$ (11) $$\rho = \chi\chi_\sigma^{-1} = \rho(\sigma),$$

(12) $$p = C + \int_0^\sigma \chi^{-2}d\sigma = p(\sigma).$$

From (4) it is clear that $\sigma = 0$ yields the sonic values q_*, ρ_*, p_*.

When σ is eliminated from (10), (12) we obtain the Bernoulli function $q = q(p)$. Equations (11) and (12) constitute parametric equations for the equation of state, the elimination of the parameter σ leading to the equation of state in the usual form $p = p(\rho)$. For a given $K(\sigma)$ the equation of state (11) (12) is uniquely determined by the choice of the initial values $\chi(0)$, $\chi'(0)$ of the solution $\chi = \chi(\sigma)$ of (9), and the constant C in (12).

We now prove the

THEOREM. *If* $K(\sigma) = \sum_{r=1}^{\infty} k_r\sigma^r$, $\bar{K}(\sigma) = \sum_{r=1}^{\infty} \bar{k}_r\sigma^r$, $k_{n+1} \neq \bar{k}_{n+1}$, $k_i = \bar{k}_i$ $(i = 1, 2, \cdots, n)$ *and if* $\chi(\sigma)$, $\bar{\chi}(\sigma)$ *are solutions of* $\chi'' - K\chi = 0$, $\bar{\chi}'' - \bar{K}\bar{\chi} = 0$, *respectively, meeting the same initial conditions* $\chi(0) = \bar{\chi}(0) = \text{const.}$, $\chi'(0) = \bar{\chi}'(0) = \text{const.}$, *then the curves*

$$C_1\colon\ \rho = \chi(\sigma)/\chi'(\sigma) = \rho(\sigma), \qquad p = p_* + \int_0^{\sigma} \chi^{-2}(\sigma)\,d\sigma = p(\sigma),$$

$$C_2\colon\ \bar{\rho} = \bar{\chi}(\sigma)/\bar{\chi}'(\sigma) = \bar{\rho}(\sigma), \qquad \bar{p} = p_* + \int_0^{\sigma} \bar{\chi}^{-2}(\sigma)\,d\sigma = \bar{p}(\sigma),$$

have contact of order [3] *at least* $n + 1$ *at the point* (ρ_*, p_*) *in the* (ρ, p)-*plane, where* $\rho_* = \rho(0) = \bar{\rho}(0)$, $p_* = p(0) = \bar{p}(0)$, *are the sonic densities and sonic pressures.*

Proof. By Maclaurin's theorem the first n derivatives of K and $\bar{K}$ agree at $\sigma = 0$ but $K^{(n+1)}(0) \neq \bar{K}^{(n+1)}(0)$. By taking the r-th derivatives of the differential equations satisfied by χ, $\bar{\chi}$ it is easy to see from Leibnitz's theorem that the $(r + 2)$-nd derivative of each of $\chi(\bar{\chi})$ can be expressed in terms of at most the r-th derivative of $K(\bar{K})$ and $\chi(\bar{\chi})$. In particular, since $\chi(0) = \bar{\chi}(0)$, $\chi'(0) = \bar{\chi}'(0)$, the differential equations assume $\chi''(0) = \bar{\chi}''(0)$ and by taking $r = 1$ we see that $\chi'''(0) = \bar{\chi}'''(0)$. Continuing in this way we see that $\chi^{(i)}(0) = \bar{\chi}^{(i)}(0)$ $(i = 0, 1, \cdots, n + 2)$.

By differentiating the equations in C_1, C_2, $n + 1$ times and evaluating for $\sigma = 0$ it follows that the $(n + 1)$-st derivaties of ρ, ρ and of p, $\bar{p}$ agree for $\sigma = 0$, to establish the theorem.

3. An approximation to a polytropic gas.

For a polytropic gas

$$\rho = kp^n, \qquad 0 < n = \gamma^{-1} < 1, \tag{13}$$

$$q^2 = \hat{q}^2 - 2p^{1-n}/k(1 - n), \tag{14}$$

where k and $\hat{q}$ denote constants for irrotational flow, $\hat{q}$ being the maximum speed. For the acoustic pressure p_* we find

$$p_* = [kn\hat{q}^2(1 - n)(1 + n)^{-1}]^{1/(1-n)} \tag{15}$$

from which we find

$$(16)\quad q_* = [(1-n)/(1+n)]^{\frac{1}{2}}\hat{q}, \qquad \rho_* = k[kn\hat{q}^2(1-n)(1+n)^{-1}]^{n(1-n)^{-1}}.$$

If $K(\sigma)$ is expanded in a power series about $\sigma = 0$,

$$(17)\qquad K(\sigma) = a\sigma - b\sigma^2 + \cdots,$$

there being no constant term due to (4) and (5). Clearly

$$a = K'(0) = K_p/\sigma_p\big|_{p=p_*} = -q^5 q_{ppp}\big|_{p=p_*}.$$

We now employ (14) to find

$$(18)\qquad a = \rho_*^{-3}(\gamma+1).$$

By suitable choice of units we can realize

$$(19)\qquad \rho_*^3 = \gamma + 1, \text{ i. e., } a = 1.$$

For the first approximation $K(\sigma) = \sigma$ to (17) the program outlined in Sections **1** and **2** for obtaining a flow in the physical plane was carried out by Martin and Thickstun [7] for the particular solution $\theta = \sigma\psi + 3^{-1}\psi^3$ of (6). It is the purpose of this paper to carry out a similar investigation beginning with the second approximation $K = \sigma - b\sigma^2$ to (17).

Accordingly, we again employ (4) and (5) to obtain

$$b = -2^{-1}K''(0) = [2^{-1}q^7 q_{pppp} - 4qq_p(-q^5 q_{ppp})]_{p=p_*}$$

which with (14), (3), (15), (16), and (19) yields

$$(20)\qquad b = 2^{-1}(2\gamma+5)(\gamma+1)^{-\frac{1}{3}}.$$

For air $\gamma = 1.4$ (approx.) and $b = 2.9129$ (approx.). It is convenient to set $\delta^2 = 2b$ and from now on we shall take

$$(21)\qquad K(\sigma) = \sigma - 2^{-1}\delta^2\sigma^2.$$

By a T_2-gas we understand a fluid whose equation of state $p = p(\rho)$ is defined by (11) and (12) for K as given in (21). Equation (9) now becomes

$$(22)\qquad \chi_{\sigma\sigma} - (\sigma - 2^{-1}\delta^2\sigma^2)\chi = 0$$

and by suitably adjusting the two arbitrary constants in the general solution of (22) and taking $C = p_*$ in (12) we can realize

$$(23)\qquad q(0) = q_*, \qquad \rho(0) = \rho_*, \qquad p(0) = p_*.$$

Thus the speed, density, and pressure of a T_2-gas along the sonic line in the physical plane can be brought into agreement with the acoustic values of these quantities for a polytropic gas.

Using (17) and (21) as K and $\bar{K}$ in the theorem of Section **2** it follows that the graphs of the equations of state for a T_2-gas and for a polytropic gas have contact of order at least three at (ρ_*, p_*). Further computation shows that the contact is exactly of order 3.

4. The equation of state for a T_2-gas. To investigate the equation of state for a T_2-gas we use (22) to study the manner in which χ varies with σ. The existence theorem for linear equations [4] assures us of a unique solution $\chi = \chi(\sigma)$, once we prescribe $\chi(0) = q_*^{-1}$ and $\chi_\sigma(0) = (\rho_* q_*)^{-1}$.

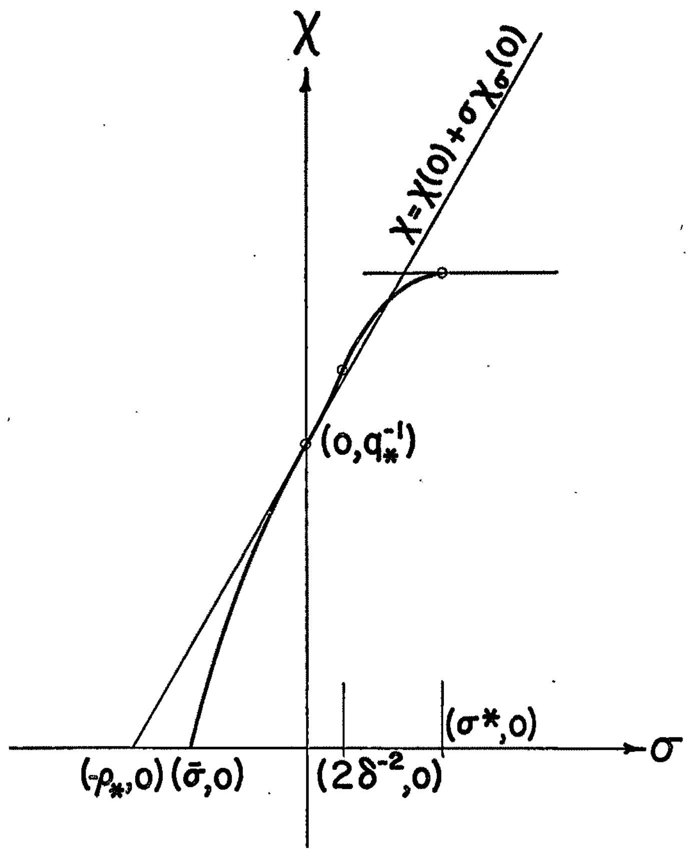

FIG. 1.

As Figure 1 indicates and (22) implies the graph of $\chi = \chi(\sigma)$ has an inflection point at $(0, q_*^{-1})$, is concave downward for $\sigma < 0$, and cuts the σ-axis at an acute angle at a point $(\bar{\sigma}, 0)$, where $-\rho_* < \bar{\sigma} < 0$. For

$0 < \sigma < 2\delta^{-2}$, the graph is concave upward with a second point of inflection at $\sigma = 2\delta^{-2}$ after which point the graph is concave downward.

The graph has a horizontal tangent for $\sigma = \sigma^*$ where $\sigma^* > 2\delta^{-2}$. To see this we write (22) in the form

$$\chi_\sigma = \chi_\sigma(2\delta^{-2}) + 2^{-1}\delta^2 \int_{2\sigma^{-2}}^{\sigma} \sigma(2\delta^{-2} - \sigma)\chi d\sigma.$$

We may assume that as σ increases $\chi > \chi(2\delta^{-2})$, otherwise χ_σ would vanish by Rolle's theorem and the assertion is immediately true. Under this assumption it is readily seen that the second member of the above equation eventually becomes negative so that $\chi_\sigma = 0$ must hold for some $\sigma = \sigma^*$ as stated.

From (8) the speed of flow is infinite for $\sigma = \bar{\sigma}$ and from (11) the density is infinite for $\sigma = \sigma^*$. We shall accordingly restrict ourselves to values of σ between $\bar{\sigma}$ and σ^*.

To study ρ as a function of σ in the interval $(\bar{\sigma}, \sigma^*)$, we note from (9) and (11) that ρ satisfies the Riccati differential equation

$$\rho' = 1 - K\rho^2 = 1 + \sigma(2^{-1}\delta^2\sigma - 1)\rho^2. \tag{24}$$

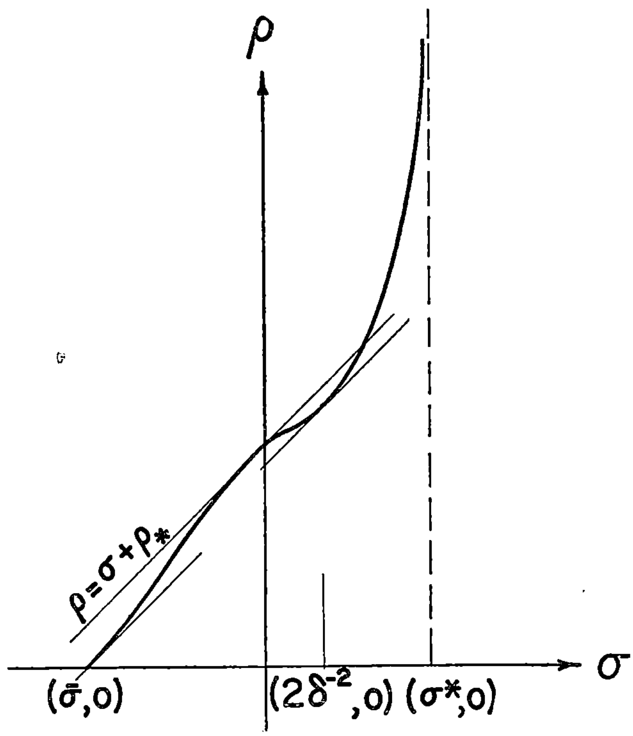

FIG. 2.

As Figure 2 indicates ρ increases monotonically from 0 to $+\infty$ as σ ranges from $\bar{\sigma}$ to σ^*. To verify that $\rho' > 0$ in the interval $(\bar{\sigma}, \sigma^*)$ we first observe from (24) that $\rho' = 1$ for $\sigma = 0$. Moreover ρ' cannot vanish in this interval. Indeed, if $\rho'(\sigma_0) = 0$ we shall have from (24)

$$\rho^2(\sigma_0) = [\sigma_0(1 - 2^{-1}\delta^2\sigma_0)]^{-1}. \tag{25}$$

This implies that $0 < \sigma_0 < 2\delta^{-2}$ and since $\rho' \leqq 1$ in the closed interval $(0, 2\delta^{-2})$ by (24), the curve $\rho = \rho(\sigma)$ lies below the straight line $\rho = \sigma + \rho_*$ in the interval under consideration. It is easy to show that the curve $\rho^2 = [\sigma(1 - 2^{-1}\delta^2\sigma]^{-1}$ lies entirely above this same line which would contradict (25). Thus ρ is an increasing function of σ in $(\bar{\sigma}, \sigma^*)$.

From (12) and (23) the pressure p is given by $p = p_* + \int_0^\sigma \chi^{-2} d\sigma$.

It is clear that p is an increasing function of σ and from Figure 1 we see that as σ tends to σ^* p tends to a finite value p_1. As σ tends to $\bar{\sigma}$, p tends to $-\infty$ since the expansion $\chi(\sigma) = a(\sigma - \bar{\sigma}) + \cdots$, $a \neq 0$, is valid about $\sigma = \bar{\sigma}$, since the graph of $\chi = \chi(\sigma)$ cuts the σ-axis at an acute angle at $(\bar{\sigma}, 0)$.

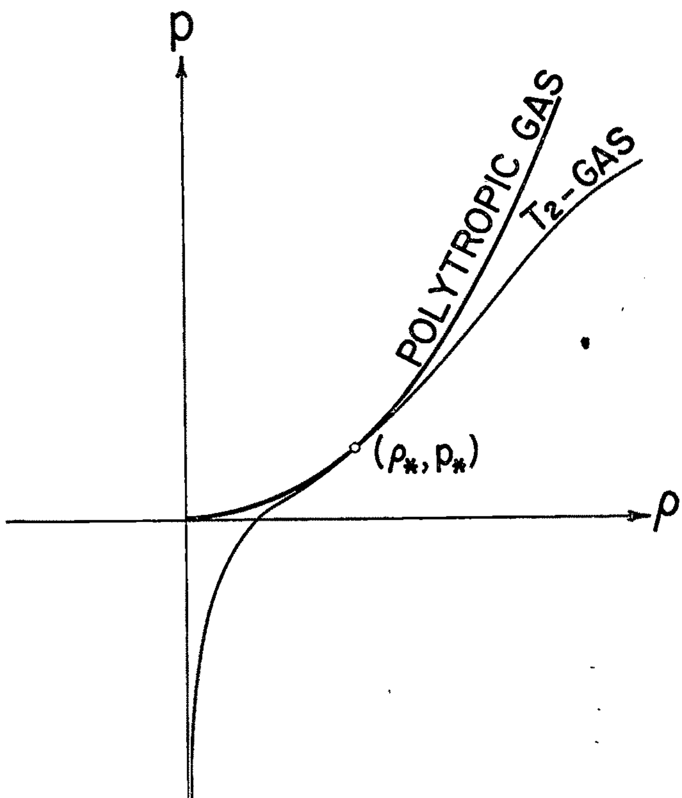

FIG. 3.

The graph $p = p(\rho)$ of the equation of state is shown in Figure 3 and is obtained by a comparison of Figures 1 and 2.

5. The direction function. We seek solutions to (6) with K defined by (21) of the form

$$\theta = \psi_0 + \psi_1\sigma + \bar{\psi}_2'\sigma^2 \tag{26}$$

where ψ_0, ψ_1, and $\bar{\psi}_2$ denote unknown functions. Substituting from (26) into (6) and integrating with respect to ψ, we find

$$\sigma - \tfrac{1}{2}\delta^2\sigma^2 = (\Sigma_0 + 2\bar{\psi}_2)(\psi_0' + \psi_1'\sigma + \bar{\psi}_2''\sigma^2) - (\psi_1 + 2\bar{\psi}_2'\sigma)^2$$

where the arbitrary function $\Sigma_0 = \Sigma_0(\sigma)$ is introduced by the integration.

We restrict ourselves to the special case $\Sigma_0 =$ constant and set

$$\Sigma_0 + 2\bar{\psi}_2 = \psi_2$$

to obtain

$$\sigma - \tfrac{1}{2}\delta^2\sigma^2 = \psi_2(\psi_0' + \psi_1'\sigma + \tfrac{1}{2}\psi_2''\sigma^2) - (\psi_1 + \psi_2'\sigma)^2$$

which, on equating coefficients of like powers of σ, yields the following system of differential equations for ψ_0, ψ_1, ψ_2.

$$\begin{aligned} &\text{(a)}\quad \psi_0'\psi_2 - \psi_1^2 = 0, \qquad \text{(b)}\quad \psi_1'\psi_2 - 2\psi_1\psi_2' = 1, \\ &\qquad\qquad \text{(c)}\quad \psi_2\psi_2'' - 2\psi_2'^2 = -\delta^2. \end{aligned} \tag{27}$$

To integrate (27c) we set $\psi_2 = \Phi^{-1}$ and this equation is replaced by $\Phi'' - \delta^2\Phi^3 = 0$ a first integral of which is $\Phi'^2 = 2^{-1}\delta^2\Phi^4 - C$, C const. If we set $C = 2^{-1}\delta^2$ we find

$$2^{-\frac{1}{2}}\delta\psi = \int_1^{\Phi} [(t^2 - 1)(t^2 + 1)]^{-\frac{1}{2}} dt, \qquad \delta > 0,$$

from which we see [9] that ψ_2 is the elliptic function $\psi_2 = cn\delta\psi$, with modulus $k = 2^{-\frac{1}{2}}$.

To find ψ_1 we multiply (27b) through by ψ_2^{-2} and integrate, to obtain $\psi_1 = \psi_2^2 \int \psi_2^{-3}\delta\psi$ which, on substituting for ψ_2 and integrating, yields $\psi_1 = \delta^{-1} sn\delta\psi dn\delta\psi$, provided we assume $\psi_1(0) = 0$.

From (27a) we find, on substituting for ψ_1, ψ_2 as given above, that $\psi_0' = 2^{-1}\delta^{-2}(nc\delta\psi - cn^3\delta\psi)$, and this, with the aid of the formulae [8]

$$\int cn^3 u du = sn u dn u; \quad \int nc u du = 2^{\frac{1}{2}} ln(2^{-\frac{1}{2}} sc u + dc u)$$

yields $\psi_0 = 2^{-\frac{1}{2}}\delta^{-3} ln(2^{-\frac{1}{2}} sc\delta\psi + dc\delta\psi) - 2^{-1}\delta^{-3} sn\delta\psi dn\delta\psi$, provided we again assume $\psi_0(0) = 0$.

On substituting for ψ_0, ψ_1, $\bar{\psi}_2$ in (26) we find the solution

$$\theta = 2^{-1}\delta^{-3}[2^{\frac{1}{2}}ln(2^{-\frac{1}{2}}sc\delta\psi + dc\delta\psi) - sn\delta\psi dn\delta\psi(1-\delta^2\sigma)^2] \tag{28}$$

to (6) for K as given in (21).

It seems unlikely that this solution could be obtained from Frankl's equation by any method which depends upon seeking for solutions explicitly of the form $\psi = \psi(\theta, \sigma)$.

It is interesting to note that if we let δ approach 0 in (21), (28), we obtain $K = \sigma$, $\theta = \sigma\psi + 3^{-1}\psi^3$, i. e. the solution to Tricomi's equation treated by Martin and Thickstun [7].

6. The mapping from the (σ, ψ)-plane to the physical plane. A computation based on (7) reveals that

$$J = \partial(x,y)/\partial(\sigma,\psi) = \theta_\psi^{-1}(K + \theta_\sigma^2)\chi\chi_\sigma,$$

which for K as in (21), and for the special solution (28), reduces to

$$J = \rho^{-1}q^{-2}cn\delta\psi. \tag{29}$$

We study the mapping upon the physical plane of the region

$$\bar{\sigma} < \sigma < \sigma^*; \quad -\delta^{-1}\kappa < \psi < \delta^{-1}\kappa, \quad \kappa = 1.85407 \text{ (approx.)}$$

where κ is the quarter-period of the elliptic function cnu with modulus $k = 2^{-\frac{1}{2}}$. It is clear from (29) that $0 < J < +\infty$ at every point of this region inasmuch as from Section **4**, ρ and q remain finite and positive for $\bar{\sigma} < \sigma < \sigma^*$. Thus the mapping upon the physical plane is one-to-one locally although a region of the physical plane may be covered more than once. The streamlines and isovels in the physical plane are the transforms of the straight lines $\psi = \text{const.}$, $\sigma = \text{const.}$ in the (σ, ψ)-plane.

The required mapping

$$z = \int q^{-1}e^{i\theta}\{cn\delta\psi d\sigma + [\delta^{-1}sn\delta\psi dn\delta\psi(1-\delta^2\sigma) + i\rho^{-1}]d\psi\} \tag{30}$$

where θ is given by (28), is obtained by substituting for θ_σ, θ_ψ from (28), for K from (21), for χ from (10), and for $\chi^{-1}\chi_\sigma$ from (11) into (7).

To obtain a streamline in the physical plane the line integral in (30) is evaluated along the path $0AP$ in Figure 4 for a fixed A and variable B; to obtain an isovel the integration is carried out along the path $0BP$, with B fixed and A variable.

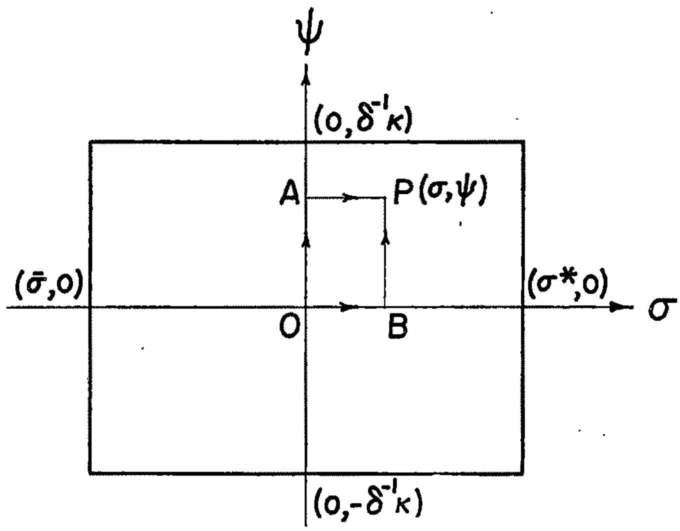

FIG. 4.

Since $z(\sigma, -\psi) = \bar{z}(\sigma, \psi)$, from (30), it is clear that the flow is symmetric with respect to the x-axis and we accordingly restrict our attention to the upper half of the physical plane.

To obtain the sonic line in the physical plane we set $\sigma = 0$ in (30), and find

$$(31) \qquad z = q_*^{-1} \int_0^{\psi} e^{i\theta} [\delta^{-1} sn\delta\psi dn\delta\psi + i\rho_*^{-1}] d\psi$$

where from (28),

$$\theta = 2^{-1}\delta^{-3}[2^{\frac{1}{2}} ln(2^{-\frac{1}{2}} sc\delta\psi + dc\delta\psi) - sn\delta\psi dn\delta\psi].$$

Thus we obtain parametric equations for the sonic line with ψ serving as parameter.

If we denote the inclination of the tangent to the sonic line by ϕ, so that $\phi = \arg z_\psi = \theta + \beta$, where

$$\beta = \text{arc cot}(\delta^{-1}\rho_* sn\delta\psi dn\delta\psi).$$

The derivatives of θ and β with respect to ψ are given by

$$\theta_\psi = 2^{-1}\delta^{-2}(nc\delta\psi - cn^3\delta\psi), \ \beta_\psi = -\delta^2\rho_* cn^3\delta\psi(\delta^2 + \rho_*^2 sn^2\delta\psi dn^2\delta\psi)^{-1}.$$

When $\psi = 0$, $\beta = \pi/2$, $\theta = 0$ and $\phi = \pi/2$. As ψ varies from 0 to $\delta^{-1}\kappa$, θ increases monotonely from 0 to $+\infty$ while β decreases monotonely from $\pi/2$ to the first quadrant angle arc-cot$(2^{-\frac{1}{2}}\delta^{-1}\rho_*)$. It follows that ϕ must decrease

until a point of inflection is reached and thereafter increases without limit. The length of the sonic line measured from the origin is given by

$$s = (\rho_* q_* \delta)^{-1} \int_0^{\psi} (\rho_*^2 sn^2 \delta t dn^2 \delta t + \delta^2)^{\frac{1}{2}} dt,$$

from which we conclude that the length of the curve cannot exceed $\kappa(\rho_* q_* \delta^2)^{-1}(2^{-1}\rho_*^2 + \delta^2)^{\frac{1}{2}}$. Hence the sonic line spirals into a finite point as is shown in Figure 5.

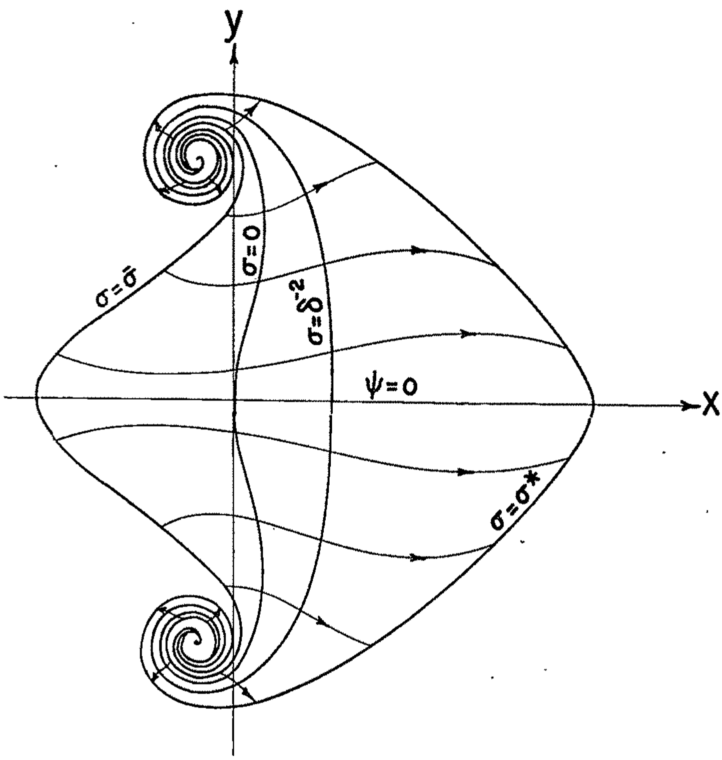

FIG. 5.

To obtain the streamline $\psi = 0$ we set $\psi = 0$ in (30) to obtain

$$x = \int_0^{\sigma} \chi(t)\, dt = x(\sigma),$$

i. e., the segment $\bar{x} < x < x^*$ of the x-axis where $\bar{x} = x(\bar{\sigma})$ and $x^* = x(\sigma^*)$ are finite points, since the corresponding areas under the curve $\chi = \chi(\sigma)$ in Figure 1 are finite.

To obtain an isovel $\sigma = \sigma_1 =$ constant we set $\sigma = \sigma_1$ in (30) and integrate along $0BP$ in Figure 4 with $0B = \sigma_1$ and A variable. This yields

$$z = x(\sigma_1) + \chi(\sigma_1) \int_0^{\psi} e^{i\theta}[\delta^{-1} sn\delta\psi dn\delta\psi(1 - \delta^2\sigma_1) + i/\rho(\sigma_1)]d\psi.$$

Then $\arg z_\psi = \theta + \beta = \phi$ where $\theta = \theta(\sigma_1, \psi)$,

$$\beta = \operatorname{arc} \cot[\delta^{-1}\rho(\sigma_1) sn\delta\psi dn\delta\psi(1 - \delta^2\sigma_1)].$$

By an argument similar to the one employed in studying the sonic line we find that all the isovels are spiral in character and intersect the x-axis at right angles. The isovel $\sigma = \delta^{-2}$, shown in Figure 5, is orthogonal to all the streamlines since $\beta = \pi/2$ at every point of this curve.

A streamline $\psi = \psi_1 =$ constant is found by setting $\psi = \psi_1$ in (30) and integrating along $0AP$ in Figure 4 with $0A = \psi_1$ and B variable. We find $z = z_1 + cn\delta\psi_1 \int_0^{\sigma} \chi e^{i\theta} d\sigma$, where θ is obtained from (28) by setting $\psi = \psi_1$ and where z_1 is the point on the sonic line corresponding to $\psi = \psi_1$.

From (28) we find that $\theta_\sigma = (2\delta)^{-1} sn\delta\psi dn\delta\psi(1 - \delta^2\sigma)$. Thus along the streamline $\psi = \psi_1$, θ increases as σ varies from $\bar{\sigma}$ to δ^{-2} and then decreases until $\sigma = \sigma^*$.

The arc length s along $\psi = \psi_1$ measured from $z = z_1$ to an arbitrary isovel $\sigma = \sigma_0$ is given by

$$s = x_0 cn\delta\psi_1, \qquad x_0 = \int_0^{\sigma_0} \chi(t)\, dt$$

where x_0 is the distance along the x-axis between the sonic line and the isovel $\sigma = \sigma_0$. It is apparent that, as ψ_1 approaches $\delta^{-1}\kappa$, the two isovels approach each other and consequently all isovels spiral into the same point.

The flow begins at the "starting line" $\sigma = \bar{\sigma}$ along which the speed is infinite. The flow particles move toward the sonic line at supersonic speed after which they move toward the isovel $\sigma = \sigma^*$ at subsonic speed upon which the density ρ becomes infinite in view of (11). The flow is illustrated in Figure 5.

Appendix.

It is interesting to compare the graphs of the functions $K = K(\sigma)$ for a T_1-gas and a T_2-gas with the graph for a polytropic gas. For a T_1-gas $K(\sigma) = \sigma$ and for a T_2-gas $K(\sigma)$ is given by (21). To graph $K = K(\sigma)$ for a polytropic gas we have from (24)

$$\rho^2 K = 1 - (d\rho/dp)/(d\sigma/dp)$$

and then employ (4), (13), and (14) to obtain $K = K(p)$, $\sigma = \sigma(p)$ as follows:

$$K = [(1+n)p^{1-n} - (1-n)kn\hat{q}^2]/[(1-n)k^2p^{1+n}], \qquad \sigma = \int_{p_*}^{p} q^{-2}dp.$$

These are parametric equations of the required curve with the pressure p serving as parameter. When $p = 0, K = -\infty$ and as p increases to p_0 (stagnation pressure) K increases monotonely from $-\infty$ to $k^{-2}p_0^{-2n}$; when $p = p_*$, $K = 0$.

Since $d\sigma/dp = q^{-2}$ we see that as p increases from $p = 0$ to $p = p_0$, σ increases monotonely from a negative constant to $+\infty$; when $p = p_*$, $\sigma = 0$.

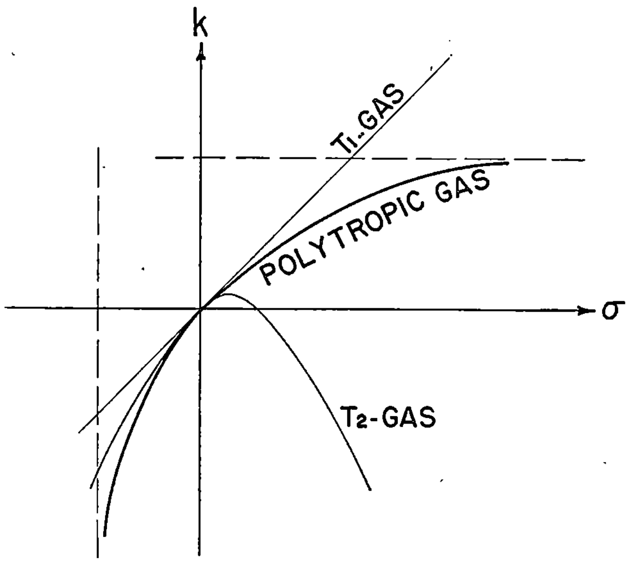

FIG. 6.

Therefore, as σ varies from a negative constant to $+\infty$, $K(\sigma)$ for a polytropic gas increases monotonely from $-\infty$ to $k^{-2}p_0^{-2n}$. The graphs of the three functions $K(\sigma)$ are exhibited in Figure 6. The shape of the curve $K = K(\sigma)$ for a polytropic gas may suggest other possible functions $K(\sigma)$ as approximations.

ANNAPOLIS, MD.

BIBLIOGRAPHY.

[1] L. Bers, "On the continuation of a potential gas flow across the sonic line," *National Advisory Committee for Aeronautics Technical Note 2058* (1950).

[2] F. I. Frankl, "On the problems of Chaplygin for mixed sub- and supersonic flows," *Translation, National Advisory Committee for Aeronautics Technical Memorandum 1155* (1947).

[3] E. Goursat, *Mathematical Analysis*, vol. 1, Translation by E, R. Hedrick, New York, 1904, pp. 443-447.

[4] E. L. Ince, *Ordinary Differential Equations*, New York, 1944, chap. 3.

[5] C. Loewner, "A transformation theory of the partial differential equations of gas dynamics," *National Advisory Committee for Aeronautics Technical Note 2065* (1950).

[6] M. H. Martin, "Steady, rotational, plane flow of a gas," *American Journal of Mathematics*, vol. 72 (1950), pp. 455-484.

[7] M. H. Martin and W. R. Thickstun, "An example of transonic flow for the Tricomi gas," *Proceedings of the Fourth Symposium in Applied Mathematics of the American Mathematical Society* (to be published).

[8] B. O. Pierce, *A short table of integrals*, New York, 1929.

[9] E. T. Whittaker and G. N. Watson, *Modern Analysis*, New York, 1948, chap. 22.

CORRECTION TO THE PAPER "ON THE SPHERICAL CONVERGENCE OF MULTIPLE FOURIER SERIES." *

By JOSEPHINE MITCHELL.

It has been pointed out by L. Schoenfeld that the proofs of Theorems 2.1 and 3.1 in [3] contain certain errors. In this note we give a proof of Theorem 2.1 which replaces the van der Corput method used in [3] by another method found in the theory of lattice points in the circle [2].

Let $e(X) \equiv e^{iX}$,

$$(1) \qquad S(R) = 4\pi^2 K_R(x,y), \qquad K_R(x,y) = 1/(4\pi^2) \sum_{m^2+n^2 \leqq R} e(mx+ny),$$

$(-\pi \leqq x, y \leqq \pi; 0 \leqq R < \infty)$ and

$$(2) \qquad L_R(\alpha,\beta) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} | K_R(\alpha - x, \beta - y) \, dx \, dy$$

$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} | K_R(x,y) | \, dx \, dy$$

(from the periodicity of the exponential system).

THEOREM 2.1 [3]. *The spherical Lebesgue function $L_R(\alpha,\beta)$ is $O(R^{1/3})$.*

Proof. The following formula may be proved as in [2, pp. 204-6]. For $R > 0$,

$$(3) \qquad \int_0^R S(t)\, dt = \sum_{j,k=-\infty}^{\infty} \int_0^R I_{jk}(t)\, dt = 4\pi R \sum_{j,k=-\infty}^{\infty} r_{jk}^{-2} J_2(r_{jk} R^{\frac{1}{2}}),$$

where

$$(4) \qquad I_{jk}(R) = \iint_{m^2+n^2 \leqq R} e(a_j m + b_k n)\, dm\, dn = 2\pi R^{\frac{1}{2}} r_{jk}^{-1} J_1(r_{jk} R^{\frac{1}{2}}),$$

$r_{jk}^2 = a_j^2 + b_k^2$, $a_j = x + 2\pi j$, $b_k = y + 2\pi k$, *and* $J_1(z)$, $J_2(z)$ are Bessel's functions. Series (3) converges absolutely. Now

* Received May 17, 1952; revised July 7, 1952.

$$(5)\qquad \int_R^{R+R^{1/3}} S(t)\,dt = 2\pi \sum_{j,k=-R^{1/6}}^{R^{1/6}} \int_R^{R+R^{1/3}} t^{\frac{1}{2}} r_{jk}^{-1} J_1(r_{jk} t^{\frac{1}{2}})\,dt$$
$$+ \{4\pi t \sum_{|j|,|k|>R^{1/6}} r_{jk}^{-2} J_2(r_{jk} t^{\frac{1}{2}})\}_{t=R}^{R+R^{1/3}}$$

$= S_1 + S_2$. From the bounds $O(z^{-\frac{1}{2}})$ for $J_1(z)$ and $J_2(z)$ and $O(j^2+k^2)$ for r_{jk}^2 if $j^2+k^2 \neq 0$, and Theorem 506 in [2] it follows that S_2 is $O(R^{2/3})$ and S_1 is $O(R^{2/3} + R^{7/12}(x^2+y^2)^{-3/4})$; $x^2+y^2 \neq 0$. Thus

$$(6)\qquad \int_R^{R+R^{1/3}} S(t)\,dt = O(R^{2/3} + R^{7/12}(x^2+y^2)^{-3/4}).$$

Also, if $R \leqq t \leqq R + R^{1/3}$, then

$$S(t) - S(R) = O(\sum_{R<m^2+n^2\leqq t} 1) = O(t - R + R^{1/3}) = O(R^{1/3})$$

([2], p. 188, (682)). Consequently

$$(7)\qquad \int_R^{R+R^{1/3}} S(t)\,dt = R^{1/3} S(R) + O(R^{2/3}),$$

so that

$$(8)\qquad S(R) = O(R^{1/3} + R^{1/4}(x^2+y^2)^{-3/4}),$$

and

$$(9)\quad L_R(\alpha,\beta) = O((4\pi^2)^{-1} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \{R^{1/3} + R^{1/4}(x^2+y^2)^{-3/4}\}\,dx\,dy) = O(R^{1/3}).$$

Remark 1. By a more precise analysis it can be proved that $L_R(\alpha,\beta)$ is $O(R^{1/4})$.

Remark 2. Unfortunately this method is not applicable to series with more than two subscripts since the series corresponding to (3) is no longer absolutely convergent.

Remark 3. From Theorem 2.1 [3], follows Theorem 2.2. of [3], namely: If

$$\sum (m^2+n^2)^{1/3} \mid a_{mn} \mid^2 < \infty, \text{ then } \lim_{R\to\infty} \sum_{\nu \leqq R} \sum_{\nu=m^2+n^2} a_{mn} e(m\alpha + n\beta)$$

exists almost everywhere on $(-\pi \leqq x \leqq \pi, -\pi \leqq y \leqq \pi)$.

However, a much stronger theorem is true for general orthogonal series as we proved in [4] by an entirely different method:

If the series $\Sigma' \log^2(m_1^2 + \cdots + m_q^2) a^2_{m_1\ldots m_q} < \infty$, then the multiple

orthogonal series $\Sigma a_{m_1 \ldots m_q} \phi_{m_1 \ldots m_q}$ is spherically convergent almost everywhere on the domain of definition of the complete orthonormal system ϕ.

From this theorem it follows easily that the R-th spherical partial sum of the multiple orthogonal series is $O(\log R)$ (cf. [1], p. 164) but apparently one can conclude nothing about the order of the spherical Lebesgue function.

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] S. Kacmarz and H. Steinhaus, *Theorie der Orthogonalreihen, Monografi Matematice*, vol. 6, Warsaw, 1935.

[2] E. Landau, *Vorlesungen über Zahlentheorie*, vol. 2, Leipzig, 1927.

[3] J. Mitchell, "On the spherical convergence of multiple Fourier series," *American Journal of Mathematics*, vol. 73 (1951), pp. 211-226.

[4] J. Mitchell, "On the spherical summability of multiple orthogonal series," *Transactions of the American Mathematical Society*, vol. 71 (1951), pp. 136-151.

COHOMOLOGY RELATIONS IN PRINCIPAL FIBER SPACES.*

By Szu-tsen Hu.

Introduction. In 1949, Eckmann published an interesting relation [1] between the Betti numbers of a finite polyhedron B and a regular covering space X of B with finite leaves. If we consider the cohomology group $H^n(X)$ with real coefficients, then $H^n(X)$ is a real vector space of finite dimension $p^n(X)$, namely, the n-th Betti number of X. If G denotes the finite group of covering transformations of X, then G acts on the vector space $H^n(X)$ by means of a linear representation of G. For each $g \varepsilon G$ let $\chi^n(g)$ denote the character of the linear transformation of $H^n(X)$ determined by g in this natural linear representation of G. Let m be the order of G. Then, according to Eckmann [1], the Betti number $p^n(B)$ of B is given by $p^n(B) = \sum_{g \varepsilon G} \chi^n(g)/m$.

The purpose of the present work is to generalize this result of Eckmann to a more general class of spaces. Since a regular covering space X over a locally connected space B is a principal fiber bundle over B with a discrete structural group G, it is natural to ask if there is an analogous relation for the compact principal fiber bundles with compact zero-dimensional structural groups. The question is answered affirmatively in the present paper in a more general form under some minor conditions. In fact, we shall prove the following theorem for the class of all compact metrizable principal fiber spaces with compact zero-dimensional structural groups in the sense of Bourbaki as defined in § 1. It obviously includes all compact metrizable principal fiber bundles with compact zero-dimensional structural groups.

Theorem I. *If X is a compact metrizable*[1] *principal fiber space over B with projection $p: X \to B$ and a compact zero-dimensional structural group G, then the homomorphism $p^*: H(B) \to H(X)$ of the Čech cohomology rings with real coefficients induced by the projection $p: X \to B$ maps $H(B)$ isomorphically onto the invariant subring $H_*(X)$ of $H(X)$.*

Here, the invariant subring $H_*(X)$ of $H(X)$ is defined as follows. By

* Received December 3, 1951; revised April 24, 1952.

[1] The metrizability which we imposed on X is used only in the construction of our third special covering in § 7. It is most likely that Theorem I is true without assuming X to be metrizable.

definition, G acts as a group of transformations on the right of X. Let $W_g: X \rightarrow X$ denote the transformation associated with the element g in G, then W_g induces an automorphism $W^*_g: H(X) \rightarrow H(X)$ of $H(X)$. The invariant subring $H_*(X)$ of $H(X)$ consists of the totality of elements held fixed by W^*_g for all elements g in G. Obviously, $H_*(X)$ is determined by $H(X)$ and the operations of G on $H(X)$ defined by $g \rightarrow W^*_g$. Hence we have the following corollary.

COROLLARY II. *Up to an isomorphism, $H(B)$ is completely determined by $H(X)$ and the operations of G on $H(X)$.*

The following corollary generalizes a theorem of G. Hirsch [2, p. 226] and B. Eckmann [1, p. 98].

COROLLARY III. *$H(B)$ is isomorphic with $H(X)$ if and only if, for each $g \varepsilon G$, W^*_g is the identity automorphism of $H(X)$.*

For each integer $n \geqq 0$, let $p^n(X)$ denote the dimension of $H^n(X)$ as a real vector space. As usual, $p^n(X)$ is called the n-th Betti number of X. Similarly we define $p^n(B)$. If, for a certain given integer $n \geqq 0$, $p^n(X)$ is finite, then $H^n(X)$ is a real vector space of finite dimension $p^n(X)$ and it can be seen that the correspondence $g \rightarrow \rho(g) = W^*_g \mid H^n(X)$ defines a linear representation ρ of degree $p^n(X)$. For each $g \varepsilon G$, let $\chi^n(g)$ denote the character of the linear transformation $\rho(g)$. According to Theorem I, $p^n(B)$ is equal to the dimension of the subspace $H_*{}^n(X)$ of $H^n(X)$ which consists of the totality of elements of $H^n(X)$ held fixed under this linear representation ρ of G. Since G is a compact group, the representation ρ is decomposable into a finite system of irreducible representations $\rho_1, \cdots \rho_r$. Then $r \geqq p^n(B)$ and exactly $p^n(B)$ of these irreducible representations are trivial representations of degree 1. If we denote by $\chi_i{}^n(g)$, $i = 1, 2, \cdots, r$, the character of the linear transformation $\rho_i(g)$, then the following equality holds, [5, p. 112], $\chi^n(g) = \chi_1{}^n(g) + \cdots + \chi_r{}^n(g)$. If ρ_i is a trivial representation of degree 1, then $\chi_i{}^n(g) = 1$ for each $g \varepsilon G$. If ρ_i is a non-trivial irreducible representation, then it is well-known [5, Theorem 24] that $\int_G \chi_i{}^n(g) = 0$. The integration is taken with respect to the unique Haar measure in G such that th measure of G is unity. Hence we obtain the following corollary which includes Eckmann's formula as a special case.

COROLLARY IV. *If $p^n(X)$ is finite, so is $p^n(B)$ and $p^n(B) = \int_G \chi^n(g) dg$.*

An application to the local properties of the homogeneous spaces is given at the end of the paper.

1. Principal fiber spaces. Let G be a topological group acting as a *group of transformations* on the *right* of a Hausdorff space X. By this we mean that, with each element g in G, there is associated a transformation $W_g: X \to X$ such that, if we use the notation $W_g(x) = xg$, the following conditions are satisfied:

(1. 1) xg is continuous in x and g simultaneously;

(1. 2) $(xg_1)g_2 = x(g_1g_2)$, $(x \varepsilon X, g_1 \varepsilon G, g_2 \varepsilon G)$;

(1. 3) $xe = x$, $(x \varepsilon X)$,

where e denotes the neutral element of G. More precisely, the condition (1. 1) means that the map $\Phi: X \times G \to X$ defined by taking $\Phi(x, g) = xg$ for each $x \varepsilon X$ and $g \varepsilon G$ is continuous. Obviously W_g is a homeomorphism of X for each $g \varepsilon G$.

Two points x and y in X are said to be *equivalent* if there exists an element g in G such that $y = xg$. This equivalence relation divides the points of X into disjoint equivalence classes called the *orbits* of G in X. The orbit which contains the point $x \varepsilon X$ will be denoted by xG. Hence $xG = yG$ if and only if x and y are equivalent. Let B denote the set of all orbits of G in X. There is a natural map $p: X \to B$ of X onto B defined by $p(x) = xG$ for each $x \varepsilon X$. Let us give B the *identification topology* determined by p. That is to say, a subset V in B is called open if and only if $p^{-1}(V)$ is an open set in X. The topological space B thus obtained will be called the *orbit space* of the transformation group G. G will be called a *regular* transformation group on the right of X if its orbit space B is a Hausdorff space.

(1. 4) LEMMA. *The natural map $p: X \to B$ is both continuous and open* [6].

Proof. The continuity of p follows directly from the definition of the identification topology in B determined by p. To prove that p is open, let U be an arbitrary open set in X and call $V = p(U)$. It remains to show that $p^{-1}(V)$ is an open set in X. By the definition of p, the set $p^{-1}(V)$ consists of the totality of the points xg in X such that $x \varepsilon U$ and $g \varepsilon G$. Hence $p^{-1}(V)$ is the union UG of the sets $W_g(U)$ for all g in G. For each g in G, W_g is a homeomorphism of X. This implies that $W_g(U)$ is open and hence, as a union of open sets, $p^{-1}(V)$ is open. This completes the proof.

In the topological product space $X \times X$, let Q denote the set which

consists of the points (x, y) in $X \times X$ such that x and y are equivalent. Q will be called the graph of the transformation group G. G is a regular transformation group on the right of X if and only if its graph Q is a closed subset of $X \times X$, [6].

If a regular transformation group G operates on the right of a Hausdorff space X in such a way that, for each point (x, y) in the graph Q of G, there exists only one element $g = u(x, y)$ in G such that $y = xg$ and that the correspondence $(x, y) \to u(x, y)$ defines a continuous map $u: Q \to G$, then X is said to be a principal *fiber space* [6] with *structural group* G. The orbits of G in X are called the *fibers* of X and the orbit space B of G is called the *base space* of X. The natural map $p: X \to B$ of X onto B will be called the *projection.* For each point x in X, the correspondence $g \to xg$ obviously defines a homeomorphism $p_x: G \to xG$ of the structural group G onto the fiber xG which contains x. p_x will be called the *perspection* at x. Obviously we have $u(x, p_x(g)) = g$ for every $x \varepsilon X$ and $g \varepsilon G$.

The class of all principal fiber spaces obviously contains all principal fiber bundles [7, p. 35]. The converse does not hold as shown by the following important example. Let X be a topological group and G a closed subgroup of X. Let G operate on the right of X by means of right translations. Then X is obviously a principal fiber space with structural group G and base space $B = X/G$, [6]. However, X is a principal fiber bundle over B if and only if there is a local cross-section [7, p. 31].

2. The cohomology rings. Throughout the sections **2-8**, we shall assume that X is a compact metrizable principal fiber space with a compact zero-dimensional structural group G. It follows that the base space B of X is also compact. We shall use all of the notations given in the previous section.

In order to apply the theorems established in a previous work [3], we shall define G to be a left transformation group of X as follows. For each element g in G, define a homeomorphism $T_g: X \to X$ of X by taking $T_g(x) = xg^{-1}$, $(x \varepsilon X)$. Then clearly we have the following properties:

(2.1) $T_g(x)$ is continuous in g and x simultaneously.

(2.2) $$T_{g_1g_2} = T_{g_1}T_{g_2}.$$

(2.3) T_e is the identity transformation of X.

Hence G is acting in this way as a transformation group on the left of X.

Throughout the paper, we shall denote by R the topological field of real

numbers and we shall use R as the only coefficient ring. Let $H(X)$ and $H(B)$ denote the *continuous cohomology rings* respectively of the spaces X and B as defined in § 4 of the previous paper [3]. Since X and B are compact Hausdorff spaces it follows from Theorem 5.1 of [3] that they are isomorphic with the Čech cohomology rings.

Let $H_I(X)$ denote the *invariant cohomology ring* [3, § 7] of X with G acting on the left of X by means of the homeomorphisms T_g. According to [3, § 7], there is a natural ring homeomorphism $j^*: H_I(X) \to H(X)$. Since G is compact, it follows from Theorem 10.1 of [3] that j^* is an isomorphism of $H_I(X)$ into $H(X)$.

For each element g in G, the homeomorphism $T_g: X \to X$ induces as in [3, § 6] an automorphism $T^*_g: H(X) \to H(X)$ of the continuous cohomology ring $H(X)$. Let $H_*(X)$ denote the subring of $H(X)$ which consists of the elements held fixed by the automorphisms T^*_g for all g in G. Obviously j^* maps $H_I(X)$ into the invariant subring $H_*(X)$ of $H(X)$. By some obvious modifications of the proof of the fundamental lemma in [3, § 9], one can prove that j^* maps $H_I(X)$ onto $H_*(X)$. Hence we have proved the following theorem.

(2.4) THEOREM. *The natural ring homomorphism $j^*: H_I(X) \to H(X)$ maps $H_I(X)$ isomorphically onto the invariant subring $H_*(X)$ of $H(X)$.*

3. The induced homomorphisms. The projection $p: X \to B$ induces according to [3, § 6] an induced homomorphism $p^*: H(B) \to H(X)$. Since, for each continuous n-function $\phi: B^{n+1} \to R$ of the base space B, the induced continuous n-function $p^\# \phi: X^{n+1} \to R$ is defined by $(p^\# \phi)(x_0, x_1, \cdots, x_n) = \phi(px_0, px_1, \cdots, px_n)$ for every $(x_0, x_1, \cdots, x_n)$ of X^{n+1} and since $pT_g = p$ for each g in G, it follows that $p^\# \phi$ is an invariant continuous n-function of X. Hence we obtain a natural induced ring homomorphism $p^*_I: H(B) \to H_I(X)$. Obviously we have

$$p^* = j^* p^*_I. \tag{3.1}$$

(3.2) The fundamental lemma. *p^*_I maps $H(B)$ isomorphically into $H_I(X)$.*

The proof of this fundamental lemma will be given in § 8.

The following theorem is a direct consequence of (3.1), (3.2), and (2.4).

(3.3) THEOREM. *The ring homomorphism $p^*: H(B) \to H(X)$ induced by the projection $p: X \to B$ maps the continuous cohomology ring $H(B)$ of B*

isomorphically onto the invariant subring $H_*(X)$ of the continuous cohomology ring $H(X)$ of X.

The Theorem I of the introduction follows from (3.3) and the natural equivalence between the Čech cohomology theory and the continuous cohomology theory of compact Hausdorff spaces.

4. Strongly invariant n-functions. Let K be an arbitrary open (and hence closed) subgroup of G. A continuous n-function $\phi: X^{n+1} \to R$ of the space X is said to be *strongly invariant with respect to K if* $\phi(x_0 g_0, \cdots, x_n g_n) = \phi(x_0, \cdots, x_n)$ for all $x_i \varepsilon X$ and all $g_i \varepsilon K$, $i = 0, \cdots, n$. If ϕ is strongly invariant with respect to K, then so is the coboundary $\delta\phi$ of ϕ.

Remark. Since the projection $p: X \to B$ is open, so is the induced map $p^{n+1}: X^{n+1} \to B^{n+1}$. Hence p^{n+1} is an identication map. Therefore by Whitehead's lemma [8], $p^\#$ gives a one-to-one correspondence between the continuous n-functions on B and those continuous n-functions on X which are strongly invariant with respect to G.

5. The strong smoothing operator I_K. For any given open subgroup K of G, we are going to define a *strong smoothing operator* I_K on the continuous n-functions of X as follows. Define a function $\kappa: G \to R$ by taking $\kappa(g) = 1$ if g is in K and $\kappa(g) = 0$ otherwise. Since K is open and closed, κ is continuous on G. Since G is a compact group, there is a unique Haar measure in G with the measure of G being 1. Let k denote the measure of K, that is to say $k = \int_G \kappa(g)\,dg$. For each continuous n-function $\phi: X^{n+1} \to R$ the continuous n-function $I_K\phi: X^{n+1} \to R$ is defined by

$$(I_K\phi)(x_0, \cdots, x_n)$$
$$= k^{-n-1} \int_G \cdots \int_G \phi(x_0 g_0, \cdots, x_n g_n)\kappa(g_0) \cdots \kappa(g_n)\,dg_0 \cdots dg_n$$

for each point $(x_0, \cdots, x_n)$ of X^{n+1}. The continuity of the n-function $I_K\phi$ follows from that of ϕ as well as the fact that xg is simultaneously continuous in x and g. The basic properties of the operator I_K are given in the following assertions.

(5.1) $$\delta(I_K\phi) = I_K(\delta\phi).$$

Proof. Let $(x_0, \cdots, x_{n+1})$ be an arbitrary point of X^{n+2}, then we have [2]

[2] The circumflex over x_i indicates that x_i is omitted, and similarly for other places.

$$(\delta I_K\phi)(x_0,\cdots,x_{n+1})=\sum_{i=0}^{n+1}(-1)^i(I_K\phi)(x_0,\cdots,\hat{x}_i,\cdots,x_{n+1})$$

$$=\sum_{i=0}^{n+1}(-1)^ik^{-n-1}\int_G\cdots\int_G\phi(x_0g_0,\cdots,\widehat{x_ig_i},\cdots,x_{n+1}g_{n+1})$$
$$\times\,\kappa(g_0)\cdots\widehat{\kappa(g_i)}\cdots\kappa(g_{n+1})dg_0\cdots\widehat{dg_i}\cdots dg_{n+1}$$

$$=\sum_{i=0}^{n+1}(-1)^ik^{-n-2}\int_G\cdots\int_G\phi(x_0g_0,\cdots,\widehat{x_ig_i},\cdots,x_{n+1}g_{n+1})$$
$$\times\,\kappa(g_0)\cdots\kappa(g_{n+1})dg_0\cdots dg_{n+1}$$

$$=k^{-n-2}\int_G\cdots\int_G\Big[\sum_{i=0}^{n+1}(-1)^i\phi(x_0g_0,\cdots,\widehat{x_ig_i},\cdots,x_{n+1}g_{n+1})\Big]$$
$$\times\,\kappa(g_0)\cdots\kappa(g_{n+1})dg_0,\cdots dg_{n+1}$$

$$=k^{-n-2}\int_G\cdots\int_G\delta\phi(x_0g_0,\cdots,x_{n+1}g_{n+1})\kappa(g_0)\cdots\kappa(g_{n+1})dg_0\cdots dg_{n+1}$$

$=(I_K\delta\phi)(x_0,\cdots,x_{n+1})$. This completes the proof of (5. 1).

(5. 2) *$I_K\phi$ is strongly invariant with respect to K.*

Proof. It follows easily from the definition of the function κ that $\kappa(hg)=\kappa(g)$, $(g\,\varepsilon\, G, h\,\varepsilon\, K)$. Then we have

$$(I_K\phi)(x_0h_0,\cdots,x_nh_n)=k^{-n-1}\int_G\cdots\int_G\phi(x_0h_0g_0,\cdots,x_nh_ng_n)\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n$$

$$=k^{-n-1}\int_G\cdots\int_G\phi(x_0h_0g_0,\cdots,x_nh_ng_n)\kappa(h_0g_0)\cdots\kappa(h_ng)_ndg_0\cdots dg_n$$

$$=k^{-n-1}\int_G\cdots\int_G\phi(x_0g_0,\cdots,x_ng_n)\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n$$

$$=(I_K\phi)(x_0,\cdots,x_n)$$

for all $x_i\,\varepsilon\, X$ and all $h_i\,\varepsilon\, K$, $i=0,\cdots,n$. This completes the proof of (5. 2).

(5. 3) *If ϕ is strongly invariant with respect to K, then $I_K\phi=\phi$.*

Proof. Let $g_0,\cdots,g_n$ be arbitrarily given elements of G. If $g_i\,\varepsilon\, K$ for all $i=0,\cdots,n$, then $\phi(x_0g_0,\cdots,x_ng_n)=\phi(x_0,\cdots,x_n)$ for every point $(x_0,\cdots,x_n)$; otherwise, there is some $g_i\not\varepsilon\, K$ and hence $\kappa(g_i)=0$. Hence, in any case, we always have

$$\phi(x_0g_0,\cdots,x_ng_n)\kappa(g_0)\cdots\kappa(g_n)=\phi(x_0,\cdots,x_n)\kappa(g_0)\cdots\kappa(g_n)$$

for all $x_i\,\varepsilon\, X$ and all $g_i\,\varepsilon\, G$, $i=0,\cdots,n$. This implies that

$$(I_K\phi)(x_0,\cdots,x_n)=k^{-n-1}\int_G\cdots\int_G\phi(x_0g_0,\cdots,x_ng_n)\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n$$
$$=k^{-n-1}\int_G\cdots\int_G\phi(x_0,\cdots,x_n)\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n$$
$$=k^{-n-1}\phi(x_0,\cdots,x_n)\int_G\cdots\int_G\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n=\phi(x_0,\cdots,x_n)$$

for every point $(x_0,\cdots,x_n)$ of X^{n+1}. This completes the proof of (5.3).

(5.4) *If K is an open normal subgroup of G and if ϕ is invariant (with respect to G), then $I_K\phi$ is also invariant (with respect to G).*

Proof. Since K is a normal subgroup of G, it follows from the definition of κ that $\kappa(gg_ig^{-1})=\kappa(g_i)$, $(g\in G, g_i\in G)$. Since ϕ is invariant and the Haar measure in a compact group is both left and right invariant, one can easily deduce the following calculations:

$$(I_K\phi)(x_0g,\cdots,x_ng)=k^{-n-1}\int_G\cdots\int_G\phi(x_0gg_0,\cdots,x_ngg_n)\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n$$
$$=k^{-n-1}\int_G\cdots\int_G\phi(x_0gg_0g^{-1},\cdots,x_ngg_ng^{-1})\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n$$
$$=k^{-n-1}\int_G\cdots\int_G\phi(x_0gg_0g^{-1},\cdots,x_ngg_ng^{-1})\kappa(gg_0g^{-1})\cdots\kappa(gg_ng^{-1})dg_0\cdots dg_n$$
$$=k^{-n-1}\int_G\cdots\int_G\phi(x_0g_0,\cdots,x_ng_n)\kappa(g_0)\cdots\kappa(g_n)dg_0\cdots dg_n=(I_K\phi)(x_0,\cdots,x_n)$$

for every point $(x_0,\cdots,x_n)$ in X^{n+1} and every element g in G. This completes the proof of (5.4).

(5.5) *If every point $x\in X$ has an open neighborhood V_x in X such that $\phi(x_0g_0,\cdots,x_ng_n)=0$ for all $x_i\in V_x$ and all $g_i\in K$, $i=0,\cdots,n$, then $I_K\phi$ is of empty support,* (see [3, § 2]).

Proof. Let x be an arbitrary given point of X. Let $(x_0,\cdots,x_n)$ be any point of X^{n+1} such that $x_i\in V_x$ for all $i=0,\cdots,n$. Consider any $n+1$ elements $g_0,\cdots,g_n$ of G. If $g_i\in K$ for all $i=0,\cdots,n$, then it follows from the hypothesis that $\phi(x_0g_0,\cdots,x_ng_n)=0$; otherwise, there is some $g_i\notin K$ and hence $\kappa(g_i)=0$. Hence, in any case, we always have

$$\phi(x_0g_0,\cdots,x_ng_n)\kappa(g_0)\cdots\kappa(g_n)=0.$$

This implies that $(I_K\phi)(x_0,\cdots,x_n)=0$ and hence x is not in the support of $I_K\phi$. Since x is arbitrary, the support of $I_K\phi$ must be empty. This completes the proof of (5.5).

6. The operator Q_K. For any given $n+1$ elements $g_0,\cdots,g_n$ of G, we construct an operator $P_{g_0,\ldots,g_n}$ which associates with each continuous $(n+1)$-

function $\phi: X^{n+2} \to R$ on X a continuous n-function $P_{g_0,\ldots,g_n}\phi: X^{n+1} \to R$ on X defined by

$$(P_{g_0,\ldots,g_n}\phi)(x_0, \cdots, x_n) = \sum_{i=0}^{n} (-1)^i \phi(x_0 g_0, \cdots, x_i g_i, x_i, \cdots, x_n)$$

for every point $(x_0, \cdots, x_n)$ in X^{n+1}.

Direct calculation shows that $\phi(x_0, \cdots, x_{n+1}) - \phi(x_0 g_0, \cdots, x_{n+1} g_{n+1})$

$$= \sum_{i=0}^{n+1} (-1)^i P_{g_0,\ldots,\hat{g}_i,\ldots,g_{n+1}} \phi(x_0, \cdots, \hat{x}_i, \cdots, x_{n+1}) + P_{g_0,\ldots,g_{n+1}} \delta\phi(x_0, \cdots, x_{n+1})$$

for any point $(x_0, \cdots, x_{n+1})$ of X^{n+2} and arbitrary elements $g_0, \cdots, g_{n+1}$ of G.

Now let K denote an open (and hence closed) subgroup with measure k and characteristic function κ as in the previous section. We construct an operator Q_K which associates with each continuous $(n+1)$-function $\phi: X^{n+2} \to R$ on X a continuous n-function $Q_K\phi: X^{n+1} \to R$ on X defined by

$$\begin{aligned}(Q_K\phi)&(x_0, \cdots, x_n) \\ &= k^{-n-1} \int_G \cdots \int_G P_{g_0,\ldots,g_n} \phi(x_0, \cdots, x_n)\kappa(g_0) \cdots \kappa(g_n) dg_0 \cdots dg_n\end{aligned}$$

for each point $(x_0, \cdots, x_n)$ in X^{n+1}. The basic properties of the operator Q_K are given in the following assertions.

(6.1) *If K is an open normal subgroup of G and if ϕ is invariant (with respect to G), then $Q_K\phi$ is also invariant (with respect to G).*

Proof. Since K is a normal subgroup of G and κ is the characteristic function of K in G, we have $\kappa(g g_i g^{-1}) = \kappa(g_i)$, $(g \in G, g_i \in G)$. Since ϕ is invariant and the Haar measure in a compact group is both left and right invariant, one can easily deduce the following calculations:

$$\begin{aligned}
(Q_K\phi)&(x_0 g, \cdots, x_n g) \\
&= k^{-n-1} \int_G \cdots \int_G P_{g_0,\ldots,g_n} \phi(x_0 g, \cdots, x_n g)\kappa(g_0) \cdots (g_n) dg_0 \cdots dg_n \\
&= k^{-n-1} \sum_{i=0}^{n} (-1)^i \int_G \cdots \int_G \phi(x_0 g g_0, \cdots, x_i g g_i, x_i g, \cdots, x_n g) \\
&\qquad\qquad \times \kappa(g_0) \cdots \kappa(g_n) dg_0 \cdots dg_n \\
&= k^{-n-1} \sum_{i=0}^{n} (-1)^i \int_G \cdots \int_G \phi(x_0 g g_0 g^{-1}, \cdots, x_i g g_i g^{-1}, x_i, \cdots, x_n) \\
&\qquad\qquad \times \kappa(g g_0 g^{-1}) \cdots \kappa(g g_n g^{-1}) dg_0 \cdots dg_n \\
&= k^{-n-1} \sum_{i=0}^{n} (-1)^i \int_G \cdots \int_G \phi(x_0 g_0, \cdots, x_i g_i, x_i, \cdots, x_n)\kappa(g_0) \cdots \kappa(g_n) dg_0 \cdots \\
&= k^{-n-1} \int_G \cdots \int_G P_{g_0,\ldots,g_n} \phi(x_0, \cdots, x_n)\kappa(g_0) \cdots \kappa(g_n) dg_0 \cdots dg_n \\
&= (Q_K\phi)(x_0, \cdots, x_n)
\end{aligned}$$

for every point $(x_0,\cdots,x_n)$ in X^{n+1} and every element g in G. This completes the proof of (6.1).

(6.2) *If every point $x \varepsilon X$ has an open neighborhood V_x in X such that $\phi(x_0g_0,\cdots,x_{n+}\,g_{n+1})=0$ for all $x_i \varepsilon V_x$ and all $g_i \varepsilon K$, $i=0,\cdots,n+1$, then $Q_K\phi$ is of empty support.*

Proof. Let x be an arbitrarily given point of X. Let $(x_0,\cdots,x_n)$ be any point of X^{n-1} such that $x_i \varepsilon V_x$ for all $i=0,\cdots,n$. Consider any $n+1$ elements $g_0,\cdots,g_n$ of G. If $g_i \varepsilon K$ for all $i=0,\cdots,n$, then it follows from the hypothesis that $\phi(x_0g_0,\cdots,x_ig_i,x_i,\cdots,x_n)=0$ for all $i=0,\cdots,n$ and hence

$$(P_{g_0,\ldots,g_n}\phi)(x_0,\cdots,x_n)=\sum_{i=0}^{n}(-1)^i\phi(x_0g_0,\cdots,x_ig_i,x_i,\cdots,x_n)=0;$$

otherwise, there is some $g_i \not\varepsilon K$ and hence $\kappa(g_i)=0$. Hence, in any case, we always have

$$P_{g_0,\ldots,g_n}\phi(x_0,\cdots,x_n)\kappa(g_0)\cdots\kappa(g_n)=0.$$

This imples that $(Q_K\phi)(x_0,\cdots,x_n)=0$ and hence x is not in the support of $Q_K\phi$. Since x is arbitrary, the support of $Q_K\phi$ must be empty. This completes the proof of (6.2).

(6.3) $\phi - I_K\phi = \delta(Q_K\phi)+Q_K(\delta\phi)$.

Proof. For any point $(x_0,\cdots,x_{n+1})$ in X^{n+2}, we have

$$\begin{aligned}
&\phi(x_0,\cdots,x_{n+1})-I_K\phi(x_0,\cdots,x_{n+1})\\
&\quad=k^{-n-2}\int_G\cdots\int_G[\phi(x_0,\cdots,x_{n+1})-\phi(x_0g_0,\cdots,x_{n+1}g_{n+1})]\\
&\qquad\qquad\times\kappa(g_0)\cdots\kappa(g_{n+1})dg_0\cdots dg_{n+1}\\
&\quad=k^{-n-2}\sum_{i=0}^{n+1}(-1)^i\int_G\cdots\int_G P_{g_0,\ldots,\hat{g}_i,\ldots,g_{n+1}}\phi(x_0,\cdots,\hat{x}_i,\cdots,x_{n+1})\\
&\qquad\qquad\times\kappa(g_0)\cdots\kappa(g_{n+1})dg_0\cdots dg_{n+1}\\
&\qquad+k^{-n-2}\int_G\cdots\int_G P_{g_0,\ldots,g_{n+1}}\,\delta\phi(x_0,\cdots,x_{n+1})\\
&\qquad\qquad\times\kappa(g_0)\cdots\kappa(g_{n+1})dg_0\cdots dg_{n+1}\\
&\quad=k^{-n-1}\sum_{i=0}^{n+1}(-1)^i\int_G\cdots\int_G P_{g_0,\ldots,g_n}\phi(x_0,\cdots,\hat{x}_i,\cdots,x_{n+1})\\
&\qquad\qquad\times\kappa(g_0)\cdots\kappa(g_{n+1})dg_0\cdots dg_{n+1}\\
&\qquad+Q_K\delta\phi(x_0,\cdots,x_{n+1})\\
&\quad=\sum_{n+1}^{i=0}(-1)^iQ_K\phi(x_0,\cdots,\hat{x}_i,\cdots,x_{n+1})+Q_K\delta\phi(x_0,\cdots,x_{n+1})\\
&\quad=\delta Q_K\phi(x_0,\cdots,x_{n+1})+Q_K\delta\phi(x_0,\cdots,x_{n+1}).
\end{aligned}$$

Since $(x_0,\cdots,x_{n+1})$ is arbitrary, this completes the proof of (6.3).

7. Some special coverings. Throughout the present section, let us denote by $\mathfrak{U} = \{U_y \mid y \in X\}$ a given covering of the space X which associates with each point $y \in X$ an open neighborhood U_y of y in X. We are going to construct a number of special refinements of $\mathfrak{U}$ which will be used in the proof of the fundamental lemma in § 8.

Since G is a compact zero-dimensional group, it follows from the simultaneous continuity of xg in x and g that, for each point $y \in X$, there exists an open neighborhood V_y of y in X and an open (and hence closed) normal subgroup K_y of G such that $xg \in U_y$, $(x \in V_y, g \in K_y)$. Since X is compact, there exist a finite number of points $y_1, \cdots, y_m$ of X such that the open sets $V_{y_1}, \cdots, V_{y_m}$ cover the space X. For simplicity of notations, let $U_i = U_{y_i}$, $V_i = V_{y_i}$, $K_i = K_{y_i}$ for each $i = 1, \cdots, m$. Thus we have constructed our *first special refinement* $\mathfrak{V} = \{V_1, \cdots, V_m\}$ of the given covering $\mathfrak{U}$.

The intersection K of the open normal subgroups $K_1, \cdots, K_m$ of G is an open (and hence closed) normal subgroup of G. K will be called *the open normal subgroup of G constructed with our first special refinement V.* The first special refinement $\mathfrak{V}$ of the covering $\mathfrak{U}$ and the corresponding open normal subgroup K of G obviously have the following property:

(7. 1) *For each open set V_i of $\mathfrak{V}$, $x \in V_i$ and $g \in K$ imply that $xg \in U_i$.*

Throughout the remainder of the section, we assume that we have a fixed first refinement $\mathfrak{V}$ of $\mathfrak{U}$ and the corresponding open normal subgroup K of G. Since K is an open subgroup of G, it follows from a lemma of J. P. Serre [6] that there exists an open covering $\mathfrak{M} = \{M_1, \cdots, M_r\}$ of the space X, called our *second special refinement* of the given covering $\mathfrak{U}$, such that it refines $\mathfrak{V}$ and satisfies the following condition:

(7. 2) *If x and y are two equivalent points* (see § 2) *in some open set of the covering $\mathfrak{M}$, then there exists an element h in K such that $y = xh$.*

So far we have not yet used the standing hypothesis that X is metrizable. To construct our third special refinement of $\mathfrak{U}$, we shall make use of the metrizability.[3] We assert that there exists a distance function $\rho : X^2 \to R$ on X which is *invariant* under G, that is to say $\rho(xg, yg) = \rho(x, y)$ $(x \in X, y \in X, g \in G)$. In fact, let $\rho^* : X^2 \to R$ be any distance function on the metrizable space X. Then an invariant distance function ρ is given by

$$\rho(x, y) = \int_G \rho^*(xg, yg)\, dg \text{ for each pair of points } x \text{ and } y \text{ in } X.$$

[3] This is the only place where the metrizability of X is used. The author has not been able to construct a refinement N of a given open covering M which satisfies (7. 3) without the help of a distance function invariant under the operations of G on X.

Let ρ be a given invariant distance function on X and denote by $\lambda(\mathfrak{M})$ the Lebesgue number of the covering $\mathfrak{M}$, [4, p. 37]. Take a finite covering $\mathfrak{N} = \{N_1, \cdots, N_s\}$ of X such that the diameter of each open set N_i of $\mathfrak{N}$ is less than $\frac{1}{2}\lambda(\mathfrak{M})$. The covering $\mathfrak{N}$ will be called our *third special refinement* of the covering $\mathfrak{U}$ and satisfies the following condition:

(7.3) *For any two open sets N_i and N_j (same or different) of the covering $\mathfrak{N}$ and any two elements g and h in G, if N_ig and N_jh have a common point, then there is an open set of covering $\mathfrak{M}$ which contains both N_ig and N_jh.*

Proof. Since the distance function ρ is invariant under G, the diameters of N_ig and N_jh are less than $\frac{1}{2}\lambda(\mathfrak{M})$. If N_ig and N_jh have a common point, then the union $N_ig \cup N_jh$ is of diameter less than $\lambda(\mathfrak{M})$. By the definition of the Lebesgue number $\lambda(\mathfrak{M})$, there is an open set of the covering $\mathfrak{M}$ which contains $N_ig \cup N_jh$. This completes the proof of (7.3).

8. Proof of the fundamental lemma. Let $n \geqq 0$ be an arbitrarily given integer. It suffices to prove that p^*_I maps the n-th continuous cohomology group $H^n(B)$ of B isomorphically onto the n-th invariant cohomology group $H^n{}_I(X)$ of X.

For convenience of the proof, let us first dispose of the simplest case $n = 0$. Let w be any element of $H^0(B)$ such that $p^*_I(w) = 0$. Choose a representative continuous 0-cocycle $c = [\phi]$ for w, where $\phi: B \to R$ is a continuous 0-function of B. Since the induced invariant continuous 0-function $p^{\#}\phi$ of X represents the element $p^*_I(w) = 0$, the support of $p^{\#}\phi$ is empty. By the definition of the support for this particular case $n = 0$, this implies that $p^{\#}\phi = 0$. Hence $\phi = 0$. This proves that p^*_I maps $H^0(B)$ isomorphically into $H^0{}_I(X)$.

To prove that p^*_I maps $H^0(B)$ onto $H^0{}_I(X)$, let z denote an arbitrary element of $H^0{}_I(X)$. z is represented by an invariant continuous 0-cocycle $d = [\psi]$ of X, where $\psi: X \to R$ is an invariant continuous 0-function of X. The invariance of ψ implies that $\psi(xg) = \psi(x)$, $(x \in X, g \in G)$; that is to say, ψ is constant on every fiber in X. Hence we obtain a single-valued map $\phi = \psi p^{-1}: B \to R$. According to a lemma of J. H. C. Whitehead [8, p. 1131], ϕ is continuous and hence a continuous 0-function of B. We are going to show that the coboundary $\delta\phi: B^2 \to R$ of ϕ has empty support. Let b be an arbitrary point in B and choose a point x in X such that $b = p(x)$. Since $[\psi]$ is a cocycle, the coboundary $\delta\psi: X^2 \to R$ must have empty support. Hence there is an open neighborhood U of x in X such that $\delta\psi(x_0, x_1)$

$= \psi(x_1) - \psi(x_0) = 0$ whenever x_0 and x_1 are in U. Call $V = p(U)$. Since p is an open map and $b = p(x)$, V is an open neighborhood of b in B. Let b_0 and b_1 be two arbitrary points in V. Choose x_0 and x_1 in U such that $b_0 = p(x_0)$ and $b_1 = p(x_1)$. Then we have $\delta\phi(b_0, b_1) = \phi(b_1) - \phi(b_0) = \psi(x_1) - \psi(x_0) = 0$. This implies that b is not in the support of $\delta\phi$. Since b is arbitrary, the support of $\delta\phi$ must be empty. Hence $[\phi]$ is a continuous 0-cocycle of B and represents an element w of $H^0(B)$. Since clearly $p^{\#}\phi = \psi$, we have $p^*{}_I(w) = z$. This proves that $p^*{}_I$ maps $H^0(B)$ onto $H^0{}_I(X)$ and disposes of the case $n = 0$.

Hereafter in the proof, we shall assume that $n > 0$. To prove that $p^*{}_I$ maps $H^n(B)$ isomorphically into $H^n{}_I(X)$, let w be any element of $H^n(B)$ such that $p^*{}_I(w) = 0$. We are going to show that $w = 0$.

w is represented by a continuous n-cocycle $c = [\phi]$ of B, where $\phi: B^{n+1} \to R$ is a continuous n-function $p^{\#}\phi: X^{n+1} \to R$ is invariant and $[p^{\#}\phi]$ is an invariant n-cocycle of X representing the element $p^*{}_I(w)$. Since $p^*{}_I(w) = 0$, there exist an invariant continuous $(n-1)$-function $\xi: X^n \to R$ and a continuous n-function $\eta: X^{n+1} \to R$ with empty support such that

$$p^{\#}\phi = \delta\xi + \eta. \tag{8.1}$$

Since both $p^{\#}\phi$ and $\delta\xi$ are invariant, so is η.

Since the support of η is empty, we may choose for each point y in X an open neighborhood U_y of y in X such that $\eta(x_0, x_1, \cdots, x_n) = 0$ whenever x_i is in U_y for each $i = 0, 1, \cdots, n$. Thus we have obtained an open covering $\mathfrak{U} = \{U_y \mid y \in X\}$ of the space X. Let $\mathfrak{V} = \{V_1, \cdots, V_m\}$ denote the first special refinement of $\mathfrak{U}$ constructed in § 7 and let K denote the corresponding open normal subgroup of G.

Let $\theta = I_K\xi: X^n \to R$, $\tau = I_K\eta: X^{n+1} \to R$. Since K is an open normal subgroup of G, it follows from (5.4) that both θ and τ are invariant. We assert that τ is of empty support. In fact, let x be an arbitrary point of X. Choose an open set V_j of the covering $\mathfrak{V}$ which contains x. By (7.1) and the definition of $\mathfrak{U}$, we have $\eta(x_0g_0, x_1g_1, \cdots, x_ng_n) = 0$ for all $x_i \in V_j$ and all $g_i \in K$, $i = 0, 1, \cdots, n$. Hence it follows from (5.5) that τ is of empty support. More precisely, we have $\tau(x_0, \cdots, x_n) = 0$ if the $n+1$ points $x_0, \cdots, x_n$ are contained in some open set V_j of the covering $\mathfrak{V}$.

Obviously $p^{\#}\phi$ is strongly invariant with respect to G (and hence K). Applying the operator I_K on both sides of (8.1) and using (5.3) and (5.1), we obtain

$$p^{\#}\phi = \delta\theta + \tau. \tag{8.2}$$

Now let $\mathfrak{M} = \{M_1, \cdots, M_r\}$, $\mathfrak{N} = \{N_1, \cdots, N_s\}$ denote respectively the second and the third special refinements of $\mathfrak{U}$ constructed in § 7. Call $L_i = p(N_i)$, $(i = 1, \cdots, s)$. Since p is an open map of X onto B, the collection $\mathfrak{L} = \{L_1, \cdots, L_s\}$ is an open covering of B. Let D denote the diagonal of the product space B^n, namely, the closed subset of B^n consisting of the points $(b_0, \cdots, b_{n-1})$ of B^n such that $b_0 = \cdots = b_{n-1}$. Define an open neighborhood L of D by means of the condition that a point $(b_0, \cdots, b_{n-1})$ is in L if and only if there is some open set L_j of the covering L which contains the n points $b_0, \cdots, b_{n-1}$.

We shall define a function $f: L \to R$ as follows. Let $(b_0, \cdots, b_{n-1})$ be an arbitrary point of L. According to the definition of L, there is some L_j containing the n points $b_0, \cdots, b_{n-1}$. Since $L_j = p(N_j)$, there are n points $x_0, \cdots, x_{n-1}$ in N_j such that $b_i = p(x_i)$ for each $i = 0, \cdots, n-1$. We define $f(b_0, \cdots, b_{n-1}) = \theta(x_0, \cdots, x_{n-1})$. To justify this definition, we have to show that the value $f(b_0, \cdots, b_{n-1})$ so obtained depends neither on the choice of L_j nor on the choice of $x_0, \cdots, x_{n-1}$ from N_j.

Let L_t be an arbitrary open set of the covering $\mathfrak{L} = \{L_1, \cdots, L_s\}$ which contains the n points $b_0, \cdots, b_{n-1}$. Choose arbitrarily n points $x^*_0, \cdots, x^*_{n-1}$ from N_t such that $b_i = p(x^*_i)$ for each $i = 0, \cdots, n-1$. Since $p(x^*_0) = b_0 = p(x_0)$, there is an element g in G such that $x^*_0 = x_0 g$. This implies that N_t and $N_j g$ have a common point and hence, by (7.3), there is an open set of the covering $\mathfrak{M} = \{M_1, \cdots, M_r\}$ which contains both N_t and $N_j g$. Since x^*_i and $x_i g$, $1 \leqq i \leqq n-1$, are equivalent points, it follows from (7.2) that there is an element k_i of K with $x^*_i = x_i g k_i$. Since θ is strongly invariant with respect to K and invariant with respect to G according to (5.2) and (5.4), we have

$$\begin{aligned}\theta(x^*_0, x^*_1, \cdots, x^*_{n-1}) &= \theta(x_0 g, x_1 g k_1, \cdots, x_{n-1} g k_{n-1}) \\ &= \theta(x_0 g, x_1 g, \cdots, x_{n-1} g) = \theta(x_0, x_1, \cdots, x_{n-1}).\end{aligned}$$

This completes the justification of the definition of the function $f: L \to R$.

We are going to show that the function $f: L \to R$ is continuous. For any open set L_j of the covering $\mathfrak{L} = \{L_1, \cdots, L_s\}$, let L_j^n denote the subset of B^n which consists of the points $(b_0, \cdots, b_{n-1})$ of B^n such that $b_i \varepsilon L_j$ for every $i = 0, \cdots, n-1$. Let N_j^n denote the subset of X^n similarly defined. L_j^n is an open set of B^n and is contained in L. Hence L_j^n is an open set of L. Let $p^n: N_j^n \to L_j^n$ denote the continuous map of N_j^n onto L_j^n defined by $p^n(x_0, \cdots, x_{n-1}) = (px_0, \cdots, px_{n-1})$ for each point $(x_0, \cdots, x_{n-1})$ of N_j^n.

The topology of L_j^n as a subspace of B^n coincides with the identification topology determined by the map p^n. According to the construction of the function $f: L \to R$, one can easily see that $f | L_j^n = \theta(p^n)^{-1}$. Hence it follows from Whitehead's lemma [8, p. 1131] that $f | L_j^n$ is continuous. Since L is the union of the open sets $L_1^n, \cdots, L_s^n$, this proves the continuity of the function $f: L \to R$.

Choose an open neighborhood L_* of the diagonal D in B^n such that the closure $Cl(L_*)$ of L_* is contained in L. According to Tietze's extension theorem, there exists a continuous function $\psi: B^n \to R$ such that $\psi(b_0, \cdots, b_{n-1}) = f(b_0, \cdots, b_{n-1})$ whenever $(b_0, \cdots, b_{n-1})$ is in $Cl(L_*)$. ψ is by definition a continuous $(n-1)$-function of B.

Let $\chi: B^{n+1} \to R$ denote the continuous n-function of B defined by $\chi = \phi - \delta\psi$. We are going to prove that the support of χ is empty. Let b be any given point of B. Since L_* is an open neighborhood of the diagonal D in B^n, there is an open neighborhood W of b in B such that W is contained in some open set L_j of the covering $\mathfrak{L} = \{L_1, \cdots, L_s\}$ and such that W^n is contained in L_*. Let $b_0, \cdots, b_n$ be any $n+1$ points in W. Since $W \subset L_j = p(N_j)$, we may choose $n+1$ points $x_0, \cdots, x_n$ of N_j such that $b_i = p(x_i)$ for each $i = 0, \cdots, n$. Then we have

$$\begin{aligned}
\chi(b_0, \cdots, b_n) &= \phi(b_0, \cdots, b_n) - \delta\psi(b_0, \cdots, b_n) \\
&= \phi(b_0, \cdots, b_n) - \sum_{i=0}^{n} (-1)^i \psi(b_0, \cdots, \hat{b}_i, \cdots, b_n) \\
&= \phi(b_0, \cdots, b_n) - \sum_{i=0}^{n} (-1)^i f(b_0, \cdots, \hat{b}_i, \cdots, b_n) \\
&= p^{\#}\phi(x_0, \cdots, x_n) - \sum_{i=0}^{n} (-1)^i \theta(x_0, \cdots, \hat{x}_i, \cdots, x_n) \\
&= p^{\#}\phi(x_0, \cdots, x_n) - \delta\theta(x_0, \cdots, x_n)
\end{aligned}$$

$= \tau(x_0, \cdots, x_n)$. By our construction of the third special refinement $\mathfrak{N} = \{N_1, \cdots, N_s\}$ in § 7, $\mathfrak{N}$ is a refinement of the first refinement $\mathfrak{V}$ and hence N_j is contained in some open set of the covering $\mathfrak{V} = \{V_1, \cdots, V_m\}$. This implies that $\tau(x_0, \cdots, x_n) = 0$. Hence we have proved that χ is of empty support.

Since $\phi = \delta\psi + \chi$ and χ is of empty support, it follows that the cocycle $[\phi]$ is a coboundary and hence $w = 0$. This completes the proof that $p^*{}_I$ maps $H^n(B)$ isomorphically into $H_I{}^n(X)$.

To prove that $p^*{}_I$ maps $H^n(B)$ onto $H_I{}^n(X)$, let z be an arbitrary element of $H_I{}^n(X)$. z is represented by an invariant n-cocycle $c = [\phi]$ of

X, where $\phi: X^{n+1} \to R$ is an invariant continuous n-function on X. Since $[\phi]$ is a cocyle, it follows that $\delta\phi$ has empty support. Hence we may choose for each point y in X an open neighborhood U_y of y in X such that $\delta\phi(x_0, x_1, \cdots, x_{n+1}) = 0$ whenever x_i is in U_y for each $i = 0, 1, \cdots, n+1$. Thus we have obtained an open covering $\mathfrak{U} = \{U_y \mid y \varepsilon X\}$ of the space X. Let $\mathfrak{V} = \{V_1, \cdots, V_m\}$ denote the first special refinement of $\mathfrak{U}$ constructed in § 7 and let K denote the corresponding open normal subgroup of G.

Consider the continuous n-function $I_K\phi: X^{n+1} \to R$. Since K is an open normal subgroup of G, it follows from (5.4) that $I_K\phi$ is invariant. By (6.3) we have

$$\phi - I_K\phi = \delta(Q_K\phi) + Q_K(\delta\phi). \tag{8.3}$$

According to (6.1), $Q_K\phi$ is an invariant continuous $(n-1)$-function on X. We assert that $Q_K(\delta\phi)$ is of empty support. In fact, let x be an arbitrary point of X. Choose an open set V_j of the covering $\mathfrak{V}$ which contains x. By (7.1) and the definition of $\mathfrak{U}$, we have $\delta\phi(x_0g_0, x_1g_1, \cdots, x_{n+1}g_{n+1}) = 0$ for all $x_i \varepsilon V_j$ and all $g_i \varepsilon K$, $i = 0, 1, \cdots, n+1$. Hence it follows from (6.2) that $Q_K(\delta\phi)$ is of empty support. Since $Q_K\phi$ is invariant and $Q_K(\delta\phi)$ is of empty support, (8.3) implies that $[I_K\phi]$ is an invariant n-cocycle of K and represents the same element z of $H_I{}^n(X)$ that $[\phi]$ does. Further, it follows from (5.1) and the proof of (5.5) that

$$\delta I_K\phi(x_0, x_1, \cdots, x_{n+1}) = I_K\delta\phi(x_0, x_1, \cdots, x_{n+1}) = 0$$

if the $n+2$ points $x_0, x_1, \cdots, x_{n+1}$ are contained in some open set V_j of the covering $\mathfrak{V}$.

Now let $\mathfrak{M} = \{M_1, \cdots, M_r\}$, $\mathfrak{N} = \{N_1, \cdots, N_s\}$ denote respectively the second and the third special refinements of $\mathfrak{U}$ constructed in § 7. Call $L_i = p(N_i)$, $(i = 1, \cdots, s)$. Then the collection $\mathfrak{L} = \{L_1, \cdots, L_s\}$ is an open covering of B. Let D denote the diagonal of the product space B^{n+1} and define an open neighborhood L of D in B^{n+1} by means of the condition that a point $(b_0, \cdots, b_n)$ is in L if and only if there is some L_j containing the $n+1$ points $b_0, \cdots, b_n$.

We shall define a function $f: L \to R$ as follows. Let $(b_0, \cdots, b_n)$ be an arbitrary point of L. According to the definition of L, there is some open set L_j of the covering $\mathfrak{L}$ which contains the $n+1$ points $b_0, \cdots, b_n$. Since $L_j = p(N_j)$, there are $n+1$ points $x_0, \cdots, x_n$ in N_j such that $b_i = p(x_i)$ for each $i = 0, \cdots, n$. We define $f(b_0, \cdots, b_n) = I_K\phi(x_0, \cdots, x_n)$. To justify this definition, one can prove by means of the method used above that the value $f(b_0, \cdots, b_n)$ depends neither on the choice of L_j nor on the choice

of $x_0, \cdots, x_n$ from N_j. One can also show that the function $f: L \to R$ thus defined is continuous.

Choose an open neighborhood L_* of the diagonal D in B^{n+1} such that the closure $Cl(L_*)$ is contained in L. According to Tietze's extension theorem, there exists a continuous function $\psi: B^{n+1} \to R$ such that $\psi(b_0, \cdots, b_n) = f(b_0, \cdots, b_n)$ whenever $(b_0, \cdots, b_n)$ is in $Cl(L_*)$. ψ is by definition a continuous n-function of B.

We are going to prove that the support of $\delta\psi$ is empty. Let b be any given point of B. Since L_* is an open neighborhood of the diagonal D in B^{n+1}, there is an open neighborhood W of b in B such that W is contained in some open set L_j of the covering $\mathfrak{L} = \{L_1, \cdots, L_s\}$ and such that W^{n+1} is contained in L_*. Let $b_0, \cdots, b_{n+1}$ be any $n+2$ points in W. Since $W \subset L_j = p(N_j)$, we may choose $n+2$ points $x_0, \cdots, x_{n+1}$ in N_j such that $b_i = p(x_i)$ for each $i = \text{J}, \cdots, n+1$. Then we have

$$\delta\psi(b_0, \cdots, b_{n+1}) = \sum_{i=0}^{n+1} (-1)^i \psi(b_0, \cdots, \hat{b}_i, \cdots, b_{n+1})$$

$$= \sum_{i=0}^{n+1} (-1)^i f(b_0, \cdots, \hat{b}_i, \cdots, b_{n+1}) = \sum_{i=0}^{n+1} (-1)^i I_K\phi(x_0, \cdots, \hat{x}_i, \cdots, x_{n+1})$$

$= \delta I_K\phi(x_0, \cdots, x_{n+1})$. By our construction of the third special refinement $\mathfrak{N} = \{N_1, \cdots, N_s\}$ in § 7, N_j is contained in some open set of the covering $\mathfrak{V} = \{V_1, \cdots, V_m\}$. This implies that $\delta I_K\phi(x_0, \cdots, x_{n+1}) = 0$. Hence $\delta\psi$ is of empty support and, therefore, $[\psi]$ is a continuous n-cocycle of B.

The continuous n-cocyle $[\psi]$ represents an element w of the n-th continuous cohomology group $H^n(B)$. We are going to prove that $p^*{}_I(w) = z$. $p^*{}_I(w)$ is represented by the invariant continuous n-cocycle $[p^\#\psi]$ of X while z is represented by $[I_K\phi]$. Hence it suffices to show that $p^\#\psi - I_K\phi$ is of empty support. Let x be any given point of X and call $b = p(x)$. Since L_* is an open neighborhood of the diagonal D in B^{n+1}, there is an open neighborhood W of b in B such that W^{n+1} is contained in L_*. Choose an open neighborhood Z of x in X such that Z is contained in some open set N_j of the covering $\mathfrak{N} = \{N_1, \cdots, N_s\}$ and such that $p(Z)$ is contained in W. Let $x_0, \cdots, x_n$ be any given $n+1$ points in Z and call $b_i = p(x_i)$ for each $i = 0, \cdots, n$. Then we have

$$p^\#\psi(x_0, \cdots, x_n) = \psi(b_0, \cdots, b_n) = f(b_0, \cdots, b_n) = I_K\phi(x_0, \cdots, x_n).$$

This proves that the point x is not in the support of $p^\#\psi - I_K\phi$. Since x is arbitrary, the support of $p^\#\psi - I_K\phi$ must be empty. Hence $p^*{}_I(w) = z$ and $p^*{}_I$ maps $H^n(B)$ onto $H_I{}^n(X)$. The proof of the fundamental lemma (3.2) is complete.

9. Applications. Throughout the present section, let Y be a locally compact metrizable topological group. Assume that $n \geqq 1$ is a given integer and that $V \subset W$ are two open neighborhoods of the neutral element e in Y such that the homomorphism $j_*: H_n(V) \to H_n(W)$ of the n-th Čech homology groups $H_n(V)$ and $H_n(W)$ (defined by means of finite open coverings and real coefficients) induced by the inclusion map $j: V \subset W$ is trivial, that is to say, J_* maps every element of $H_n(V)$ into the zero element of $H_n(W)$. Choose a compact neighborhood K of e in Y such that $KK \subset V$.

Let G be a zero-dimensional compact subgroup of Y contained in K. Denote by $M = Y/G$ the homogeneous space of the left cosets of G in Y and $\xi: Y \to M$ the natural map of Y onto M. According to an assertion of J. P. Serre [6], Y is a principal fiber space over M with structural group G.

Call $B = \xi(K)$, $X = \xi^{-1}(B)$, $p = \xi \mid X$. Then X is a compact metrizable principal fiber space over B with projection $p: X \to B$ and structural group G which is zero-dimensional. According to Theorem I in the introduction, the induced homomorphism $p^*: H^n(B) \to H^n(X)$ of the Čech homology groups with real coefficients is an isomorphism into. It follows from the duality theorem between homology and cohomology with coefficients in a field [4] that the induced homomorphism $p_*: H_n(X) \to H_n(B)$ of the Čech homology groups with real coefficients is onto.

Now let $D = \xi(W)$. Then D is an open neighborhood of $\xi(e)$ in M. Obviously, $X = KG \subset V \subset W$ and $B \subset D$. Call $q = \xi \mid W$ and consider the following diagram of continuous maps

$$\begin{tikzcd} X \arrow[r, "i"] \arrow[d, "p"] & V \arrow[r, "j"] & W \arrow[d, "q"] \\ B \arrow[rr, "k"] & & D \end{tikzcd}$$

where i, j, k are inclusion maps. This diagram gives rise to a corresponding diagram of the induced homomorphisms of the Čech homology groups with real coefficients:

$$\begin{tikzcd} H_n(X) \arrow[r, "i_*"] \arrow[d, "p_*"] & H_n(V) \arrow[r, "j_*"] & H_n(W) \arrow[d, "q_*"] \\ H_n(B) \arrow[rr, "k_*"] & & H_n(D) \end{tikzcd}$$

Clearly we have $kp = qji$. Hence it follows that $k_*p_* = q_*j_*i_*$. Since p_* is onto and j_* is trivial, this implies that k_* is also trivial. We have proved the following theorem.

(9.1) THEOREM. *The homomorphism $k_*: H_n(B) \to H_n(D)$ induced by the inclusion map $k: B \subset D$ is trivial, that is to say, k_* maps every element of $H_n(B)$ into the zero element of $H_n(D)$.*

In terms of cycles, (9.1) states that every Čech n-cycle with real coefficients in B is homologous to zero in D.

To clarify the significance of (9.1), we shall introduce a new notion as follows. Let D be a given open neighborhood of a point m in a topological space M. M is said to be *semi-locally n-connected at m relative to D with respect to real coefficients* if there is a closed neighborhood B of m such that every Čech n-cycle with real coefficients in B is homologous to zero in D.

Assume Y to be a locally compact and locally shrinkable metrizable topological group. Let W be a given neighborhood of the neutral element e in Y. Since Y is locally compact and locally shrinkable, there is an open neighborhood V of e in Y such that the closure $Cl(V)$ is compact and shrinkable in W to a point. Choose a compact neighborhood K of e in Y such that $KK \subset V$. Then, for every integer $n \geqq 1$, the homomorphism $j_*: H_n(V) \to H_n(W)$ induced by the inclusion map $j: V \subset W$ is trivial. Hence (9.1) gives the following theorem.

(9.2) THEOREM. *If G is a compact zero-dimensional subgroup of Y contained in K, then the homogeneous space $M = Y/G$ with the natural map $\xi: Y \to M$ is semi-logically n-connected at $\xi(e)$ relative to $\xi(W)$ with respect to real coefficients for every $n \geqq 1$.*

THE INSTITUTE FOR ADVANCED STUDY AND
TULANE UNIVERSITY.

BIBLIOGRAPHY.

[1] B. Eckmann, "Coverings and Betti numbers," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 95-101.

[2] G. Hirsch, "Sur les groupes d'homologie de certains complexes de recouvrement," *Potugaliae Math.*, vol. 4 (1945), pp. 225-237.

[3] S. T. Hu, "Cohomology rings of compact connected groups and their homogeneous spaces," *Annals of Mathematics*, vol. 55 (1952), pp. 391-419.

[4] S. Lefschetz, "Algebraic topology," *American Mathematical Society Coll. Publ.*, vol. 27 (1942).

[5] L. Pontrjagin, *Topological Groups*, Princeton Mathematical Series, No. 2 (1939).

[6] J. P. Serre, *Compacité locale des espaces fibrés*, C. R. Acad., Paris, vol. 229 (1949), pp. 1295-1297.

[7] N. E. Steenrod, *The Topology of Fibre Bundles*, Princeton Mathematical Series, No. 14 (1951).

[8] J. H. C. Whitehead, "Note on a theorem due to Borsuk," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 1125-1132.

LINEARLY COMPACT MODULES AND RINGS.*

By DANIEL ZELINSKY.

In [10, Theorem 7] some device besides countability was needed that would guarantee that an inverse limit was not zero unless the individual terms were zero. As noted there (added in proof), Dieudonné suggested that in the case of vector spaces the issue can be settled by linear compactness. This led me to generalize to topological modules over general rings this concept of linear compactness introduced by Lefschetz [9] and exploited recently by Dieudonné ([3] and [4]). The generalization is made in the obvious way (section 1) and is quite general, though the modules that actually come up in this paper are only rings considered as modules over themselves, and submodules of these. We might also note that all rings considered are topological rings with ideal neighborhoods of zero (e. g., rings with the discrete topology). Linear compactness in general seems to work best in the presence of submodule neighborhoods of zero.

Section **1** is occupied with the statements (and proofs in the nontrivial cases) of analogs of theorems about compact spaces that remain valid for linearly compact modules. Section **2** applies these results to eliminate the undesirable countability hypotheses from the theorem of [10] referred to above, and to weaken the other hypotheses of that theorem as well. The final result (Theorem 1) seems to be the natural generalization of Kaplansky's decomposition theorem for compact semisimple rings [7, Theorem 1], essentially a proof that Kaplansky's compactness hypothesis can be replaced by linear compactness without paying any price other than allowing the direct summands in the decomposition to be rings with minimum condition instead of finite rings.

Section **3** shows that linear compactness is fruitful even in the other kind of decomposition theorem considered in [10, Theorem 5]. Commutative rings also decompose in the presence of linear compactness into primary summands and a radical ring. This theorem (Theorem 2) is again a natural generalization of the results of Kaplansky [7, Theorem 17] and van Dantzig [2, TR 39], both of these last results being concerned with rings that are linearly compact and satisfy stronger hypotheses as well.

The primary rings which occur as summands in Theorem 2 are inverse

* Received March 20, 1952.

limits of discrete, linearly compact primary rings with units. These latter might be considered the building blocks for linearly compact commutative rings with units, but exactly what kinds of rings they can be is not yet clear. This class of rings is discussed to a certain extent in Section **4**, where we remark that the class contains all fields, all maximal valuation rings, and all complete local rings. This class may be of further interest in view of results of Kaplansky [1] which extend Lefschetz's duality theory of linearly compact vector spaces to an analogous duality theory of linearly compact modules over a complete discrete valuation ring (note that by our remark above, the rings of coefficients in both Lefschetz's and Kaplansky's cases are linearly compact in the discrete topolgy).

1. Definition and fundamental properties. This paper is concerned with topological R-modules (modules M over the topological ring R [1, pp. 1 ff.] which are topological groups and in which the module product $(r, m) \rightarrow rm$ is a continuous function from $R \times M$ into M). We shall consider only left R-modules, though of course all results apply equally well to right-modules. We shall often be interested in the case where R or M or both carry the discrete topology, in which case all topological assumptions may be omitted. We shall use the phrase *linear variety* in M to mean a coset of a submodule of M. Now we may state the basic definition:

Definition. A topological R-module M is linearly compact in case every collection of closed linear varieties in M with the finite intersection property has a nonvoid intersection. A topological ring R is linearly compact in case it is a linearly compact (left) R-module.

An elementary remark that we shall use repeatedly is this: If M is an R-module and A is a closed ideal in R with $AM = 0$, then M is linearly compact as an R-module if and only if it is linearly compact as an R/A-module (the closed R-linear varieties coincide with the closed R/A-linear varieties). In particular if a topological factor ring of a ring R is a linearly compact R-module, it is a linearly compact ring.

Many of the standard theorems on compact spaces and linearly compact vector spaces can be generalized to the case of linearly compact modules. The following propositions are such generalizations; we state them without proof whenever the proof is obvious or effectively identical with that given in the reference cited.

[1] Oral communication.

PROPOSITION 1. *A cartesian product (complete direct sum) of linearly compact R-modules is a linearly compact R-module.* [9, (II, 27.2) and (I, 24.1)]

PROPOSITION 2. *If M is a linearly compact R-module and f is a continuous (module-) homomorphism of M into another topological R-module, then $f(M)$ is linearly compact.* [9, (II, 27.4)]

Remark. The mapping $f\colon M \to f(M)$ is not necessarily open unless R is restricted somehow (e.g., to be a field, or a complete discrete valuation ring with the valuation topology; cf. Section **4**).

PROPOSITION 3. *If M is a linearly compact R-module and N a closed submodule of M, then N is linearly compact.* [Obvious]

The next proposition concerns an inverse system of R-modules, which is defined in the natural way: a collection $\{M_\alpha\}$ of topological R-modules indexed by a directed system $\{\alpha\}$, together with continuous module-homomorphisms $\pi_{\alpha\beta}\colon M_\alpha \to M_\beta$ for every pair of indices with $\alpha > \beta$. These π's are also assumed to satisfy $\pi_{\alpha\beta}\pi_{\beta\gamma} = \pi_{\alpha\gamma}$ for all $\alpha > \beta > \gamma$. The inverse limit is defined as the following topological submodule of the cartesian product of the M's: $\{x = \{x_\alpha\} \mid \pi_{\alpha\beta}x_\alpha = x_\beta \text{ for all } \alpha > \beta\}$. The natural projections π_β of the limit into the M_β are defined by $\pi_\beta(\{x_\alpha\}) = x_\beta$.

PROPOSITION 4. *An inverse limit of linearly compact R-modules is linearly compact. Each natural projection π_α is onto if all the homomorphisms $\pi_{\alpha\beta}$ are onto.* [9, (II, 27.6) and (I, 38, 39)]

PROPOSITION 5. *If M is a topological R-module with minimum condition on closed submodules, then M is linearly compact.*

Proof. Let $\{F_\alpha\}$ be a collection of closed linear varieties with the finite intersection property, and let $\{V_\alpha\}$ be the corresponding submodules of which the F_α's are cosets. A finite intersection $\bigcap_i F_{\alpha_i}$ is a coset of $\bigcap_i V_{\alpha_i}$. Let V be a submodule minimal among these finite intersections of the V_α's—say $V = \bigcap V_{\alpha_i}$—and let $F = \bigcap F_{\alpha_i}$. Then for every F_α, $F_\alpha \cap F$ is a coset of $V_\alpha \cap V$; the latter equals V because V is minimal; hence $F_\alpha \cap F$ is a coset of V contained in the coset F; thus $F_\alpha \cap F = F$, every F_α contains F, and $\bigcap_\alpha F_\alpha$ is nonvoid.

As a corollary of Proposition 5 we may assert that any module (over a discrete ring) with minimum condition on submodules is linearly compact in the discrete topology. This is an analog of the theorem [9, (II, 27.7)]

that a finite dimensional linear space is linearly compact in the discrete topology. (Also, the discrete topology is the only topology a finite dimensional linear space can carry if it has subspace neighborhoods of zero; the analogous remark is also true for modules with minimum condition.) However the converse (every linearly compact, discrete vector space is finite dimensional) does not generalize in any obvious way (cf. Theorems 3 and 4).

The best approximate converse we can get at present—which will suffice for our purposes—is a result expressed in terms of independent submodules. We shall call a collection of submodules independent in the module M in case every finite subcollection $\{M_1, \cdots, M_n\}$ satisfies $M_i + \bigcap_{j \neq i} M_j = M$ $(i = 1, \cdots, n)$. This condition is equivalent to each of the following statements [1, § 1, no. 7]: (α) The mapping $x \to (x + M_1, \cdots, x + M_n)$ induces an isomorphism of $M / \bigcap_i M_i$ onto the direct sum $\sum_{i=1}^{n} (M/M_i)$. (β) For every n elements $x_1, \cdots, x_n$ in M, $\bigcap_i (x_i + M_i)$ is not void.

Proposition 6. *Let M be a linearly compact R-module with the discrete topology. Then every independent collection of submodules in M is finite.*

Proof. Let $\{M_\alpha\}$ be an independent collection of submodules, and $\{x_\alpha\}$ any elements of M. By the independence condition (β), the linear varieties $x_\alpha + M_\alpha$ have the finite intersection property. Since M is linearly compact, there is an element in $\bigcap_\alpha (x_\alpha + M_\alpha)$.

Next define $M_0 = \{x \in M \mid x \in M_\alpha$ for all but a finite number of α's$\}$. M_0 is a submodule of M. Then for each α choose $x_\alpha \notin M_\alpha$ and consider all the linear varieties $x_\alpha + M_\alpha$ and M_0. These have the finite intersection property because as just remarked we can find an element x which is simultaneously in $x_\alpha + M_\alpha$ for a finite number of α's and in $0 + M_\alpha$ for the remaining α's (thus $x \in M_0$). Once again by linear compactness, there is an element x_0 in M_0 and in all $x_\alpha + M_\alpha$. Since $x_\alpha \notin M_\alpha$ the coset $x_\alpha + M_\alpha$ does not meet M_α, and so $x_0 \notin M_\alpha$ for every α. But x_0 is in all but a finite number of M_α's since $x_0 \in M_0$. Hence there is only a finite number of M_α's.

The following three results can be proved only when the topological module has an open base at zero consisting of submodules.

Proposition 7. *Let M be a topological R-module with submodule neighborhoods of zero. Then every linearly compact submodule of M is closed.* [9, (II, 27. 5)]

PROPOSITION 8. *A linearly compact R-module with submodule neighborhoods of zero is complete.* [3, p. 14]

PROPOSITION 9. *If M is a topological R-module with submodule neighborhoods of zero, N a closed submodule, and if N and M/N are linearly compact, then M is linearly compact* (cf. [9, (II, 5. 5)]).

Proof. Let h be the natural homomorphism of M onto M/N and $\{F_\alpha\}$ a collection of closed linear varieties in M with the finite intersection property. Without loss of generality we may assume the collection is closed under finite intersections. Then $\{\overline{h(F_\alpha)}\}$ is a collection of closed linear varieties in M/N with the finite intersection property. Thus there is some $h(x)$ in M/N such that each $h(F_\alpha)$ meets every neighborhood of $h(x)$. That is, for every submodule neighborhood V of zero in M, $[h(x) + h(V)] \cap h(F_\alpha)$ is nonvoid; taking inverse images in M, $[(x + V + N)] \cap (F_\alpha + N)$ is nonvoid; that is, $F_\alpha - x + V$ meets N.

We now show $F_\alpha - x$ meets N for each α. If F_α is a coset of the submodule V_α then $F_\alpha - x + V$ is a coset of the open (hence closed) submodule $V_\alpha + V$. Hence $N(V) = (F_\alpha - x + V) \cap N$ is a nonvoid, closed linear variety in N. For fixed α and variable V, these $N(V)$ have the finite intersection property: $N(V_1) \cap \cdots \cap N(V_n) \supset N(V_1 \cap \cdots \cap V_n)$, which is not empty. Hence $\bigcap_V (N) V$ is not empty. But

$$\bigcap_V N(V) = [\bigcap_V (F_\alpha - x + V)] \cap N = (F_\alpha - x) \cap N$$

because the intersection of all neighborhoods $F_\alpha - x + V$ of $F_\alpha - x$ is the closure of $F_\alpha - x$ (true in any topological group) and $F_\alpha - x$, like F_α, is closed. Hence $F_\alpha - x$ meets N.

Now consider the sets $(F_\alpha - x) \cap N$, which are nonvoid closed linear varieties in N. Since the collection $\{F_\alpha\}$ is closed under finite intersection, the collection $\{(F_\alpha - x) \cap N\}$ is also, and hence has the finite intersection property. Since N is linearly compact, there is an element y of N in every $F_\alpha - x$. Then $x + y$ is in every F_α, proving the proposition.

2. Linearly compact semisimple rings. In this section and the next, we intend to study a topological ring R with ideal neighborhoods of zero, considered as a left R-module, and investigate the consequences of the linear compactness of this module. Since properties of rings with ideal neighborhoods of zero are determined by the discrete homomorphs of the ring, we study the case of discrete rings first.

PROPOSITION 10. *If R is a primitive ring which is linearly compact in the discrete topology, then R is a simple ring with minimum condition: the ring of all linear transformations on a finite dimensional vector space over a division ring.*

Proof. By [5, Theorem 23] R is a dense ring of linear transformations on a vector space V over a division ring D. It suffices to show that V is finite dimensional. Consider any linearly independent set W in V, and for each elements w in W consider its annihilator in R: $\{x \varepsilon R \mid x(w) = 0\}$. We shall show that these annihilators are independent submodules of R. In fact, let $z \varepsilon R$, let $w_1, \cdots, w_n$ be a finite subset of W and let $M_1, \cdots, M_n$ be the corresponding annihilators. If z is any element of R, then by the density of R we can find an x such that $x(w_i) = 0$ and $x(w_j) = z(w_j)$ for $j \neq i$. Then $x \varepsilon M_i$, $z - x \varepsilon M_j$ for $j \neq i$, $z = x + (z - x) \varepsilon M_i + \bigcap_{j \neq i} M_j$, which proves the independence of the M's. But by Proposition 6, there can be at most a finite number of independent submodules, and so the elements of W can have at most a finite number of different annihilators. Due to the density of R, distinct elements of V have distinct annihilators; hence W is finite and V is finite-dimensional.

Before attacking the discrete, semisimple case, we shall prove two elementary lemmas that are applicable to all rings, even non-associative ones.

LEMMA 1. *Let $R = \sum A_i$ be the direct sum of a finite number of rings A_i, each of which is equal to its square: $A_i^2 = A_i$. Let $A_i' = \sum_{j \neq i} A_j$. Then the only ideal in R relatively prime to all the A_i' is R itself.*

Proof. Let A be such an ideal. Since $A + A_i' = R$, the natural projection of R onto the i-th component A_i carries A onto all of A_i. Hence $AA_i = A_i^2 = A_i$; $A_i = AA_i \subset A$ for all i; $R \subset A$.

LEMMA 2. *Let R be a ring, $B_1, \cdots, B_n$ a finite set of distinct maximal ideals in R, with each R/B_i not a zero ring. Then the ideals $B_1, \cdots, B_n$ are independent (both one- and two-sided) R-modules and $R/(B_1 \cap \cdots \cap B_n) \cong R/B_1 \oplus \cdots \oplus R/B_n$ (ring direct sum, natural isomorphism).*

Proof. For $n = 1$, the lemma is trivial. Suppose the lemma is established for $n = k$ and let $B_1, \cdots, B_{k+1}$ be maximal ideals with $(R/B_i)^2 \neq 0$; then $(R/B_i)^2 = R/B_i$. By symmetry, to prove independence of the B's it suffices to show that $B_{k+1} + \bigcap_{i=1}^{k} B_i = R$. Consider the ring

$$\bar{R} = R/(B_1 \cap \cdots \cap B_k);$$

it satisfies the conditions of Lemma 1 with $A_i = R/B_i$ and with $A_i' =$ the image of B_i in $\bar{R}$. Let A be the image of B_{k+1} in $\bar{R}$. Since $B_{k+1} + B_i = R$, we have $A + A_i' = \bar{R}$ for $i = 1, \cdots, k$, and so $A = \bar{R}$ by Lemma 1. Taking the inverse image of A in R, we get $B_{k+1} + \bigcap_{i=1}^{k} B_i = R$, as desired. Then $x \to (x + B_1, \cdots, x + B_{k+1})$ is a homomorphism of R onto $R/B_1 \oplus \cdots \oplus R/B_{k+1}$ with kernel $B_1 \cap \cdots \cap B_{k+1}$, completing the proof of the lemma.

PROPOSITION 11. *If R is a semisimple ring which is linearly compact in the discrete topology, then R is a semisimple ring with minimum condition: a finite direct sum of matrix rings over division rings.*

Proof. By [5, Theorem 25] there is a set $\{B_\alpha\}$ of ideals in R with zero intersection and with R/B_α primitive. Of course, each R/B_α is still a discrete and linearly compact R-module by Proposition 2, hence also a linearly compact ring. Thus each R/B_α satisfies the hypotheses of Proposition 10 so that each B_α is a maximal ideal with R/B_α not a zero ring. By Lemma 2, the B's are independent submodules of R, and so are finite in number (Proposition 6). The second part of Lemma 2 then shows that R is isomorphic to the direct sum of the (finite number of) classical simple rings R/B_α, proving Proposition 11.

The apparatus of [10] is now available for proving a decomposition theorem for linearly compact semisimple rings with ideal neighborhoods of zero. Only one more property of rings with minimum condition must be extended to linearly compact rings. This we proceed to do:

PROPOSITION 12. *Let R be a linearly compact, discrete ring with radical N, and f a homomorphism of R into another ring. Then the radical of $f(R)$ is exactly $f(N)$.*

Proof. $f(R)/f(N)$ is a homomorph of R/N, which is a semisimple ring with minimum condition by Proposition 11. Hence $f(R)/f(N)$ is also semisimple (and satisfies a minimum condition). Therefore $f(N)$ contains the radical of $f(R)$. The opposite inclusion is true in general: the radical of $f(R)$ always contains $f(N)$.

THEOREM 1. *Let R be a semisimple linearly compact ring with ideal neighborhoods of zero. Then R is algebraically and topologically isomorphic to a complete direct sum of discrete simple rings with minimum condition (i. e., of matrix rings over division rings).*

Proof. Let $\{V_\alpha\}$ be the ideal neighborhoods of zero in R and $R_\alpha = R/V_\alpha$

the discrete homomorphs. The R_α naturally form a "minimal" inverse system (each π_α is onto, in the notation of Section **1**), and, since R is complete (Proposition 8), the inverse limit of this inverse system is R [10, Theorem 3]. Moreover, if N_α is the radical of R_α, then the N_α's also form an inverse system which is "onto" (each $\pi_{\alpha\beta}$ is onto, in the notation of Section **1**) because of Proposition 12. The inverse limit of this system of N_α's is equal to the radical of R [10, Lemma 5; note a misprint in the statement of this lemma: for "Ψ is minimal" read "Ω is minimal"]. Of course, each N_α is a linearly compact R-module, so by Proposition 4 the natural projection of the limit into each N_α is onto.[2] But the radical of R is zero; hence every one of its projections is zero; every N_α is zero; each R_α is a linearly compact discrete, semisimple ring which must be semisimple with minimum condition. Then the R_α's decompose concordantly and the proof is completed exactly as in the proof of [10, Theorem 7].

Theorem 1 not only eliminates the countability hypothesis of [10, Theorem 7] but formally generalizes that theorem since for topological rings with ideal neighborhoods of zero, completeness together with a minimum condition modulo open ideals imply linear compactness (Proposition 5, [10, Theorem 3], and Proposition 4). Of course as a corollary of Theorem 1 we have the interesting result that completeness plus this restricted minimum condition is actually equivalent to linear compactness in the presence of semisimplicity.

3. Linearly compact commutative rings.

Once again we intend to imitate the proof of the decomposition theorem [10, Theorem 5]. A glance through this proof shows that the only property of rings with minimum condition used is: every decomposition is finite; and orthogonal idempotents modulo the radical can be raised to orthogonal idempotents in the ring. We proceed to establish this latter result for linearly compact discrete rings.

PROPOSITION 13. *Let R be a commutative ring linearly compact in the discrete topology, and let N be its radical. If $\bar{e}$ is an idempotent in R/N, there is an idempoetnt e in R whose residue class modulo N is $\bar{e}$.*

Proof. Consider the class of all pairs (f, A) with $f \varepsilon R$, the residue class of f modulo N equal to $\bar{e}$, A an ideal of R contained in N, and $f^2 - f \varepsilon A$. Partially order the class by defining $(f, A) > (f', A')$ in case $A \subset A'$ and $f \equiv f' \pmod{A'}$. Every chain has an upper bound because of the linear

[2] We obtain an interesting by-product: The radical of a linearly compact ring with ideal neighborhoods of zero is closed.

compactness of R as an R-module. Hence there is a maximal pair (e, B). We shall show $e^2 = e$.

For the sake of clarity, we complete the proof assuming that R has an identity element; if this is not the case, standard methods of replacing inverses by quasi-inverses will carry the proof through. Since $(1 - 2e)^2 = 1 + 4(e^2 - e)$ and $e^2 - e$ is in the radical, $(1 - 2e)^2$ has an inverse; thus $1 - 2e$ also has an inverse. Let $y = (e^2 - e)(1 - 2e)^{-1}$. Then $y \varepsilon B$. Compute $(e + y)^2 - (e + y) = e^2 - e + (2e - 1)y + y^2 = y^2$. If we let $f = e + y$ and A be the ideal generated by y^2, we have $(f, A) > (e, B)$. By the maximality of (e, B), we have $A = B$, so that $y \varepsilon A$, $y = ny^2 + ry^2$ with $r \varepsilon R$ and n an integer. Thus $y(1 - ny - ry) = 0$; but $ny - ry \varepsilon B \subset N$, so $1 - ny - ry$ has an inverse, and $y = 0$. Recalling that $(e + y)^2 - (e + y) = y^2$, we have $e^2 = e$, as desired.

We are now in a position to take over the proof used in [10, Section 4] for commutative rings. First, let R be a commutative ring which is linearly compact in the discrete topology and let N be its radical. Then R/N is a direct sum of a finite number of fields (Proposition 11) whose identity elements are orthogonal idempotents $\bar{e}_1, \cdots, \bar{e}_n$. By Proposition 13, these idempotents may be raised to idempotents $e_1, \cdots, e_n$ of R. Note that these e_i are uniquely determined [10, Lemma 2] and that they are orthogonal (e_ie_j is idempotent and in the radical, hence is zero). R is decomposable into the direct sum of the Re_i $(i = 1, \cdots, n)$ and S, the intersection of the annihilators of the e_i. Each Re_i is a primary ring with unit, and S is in the radical of R. Moreover, due to the uniqueness of the e_i, this decomposition is unique. This proves

PROPOSITION 14. *If R is a commutative ring which is linearly compact in the discrete topology then R is uniquely decomposable as a direct sum of a radical ring and a finite number of primary rings with units.*

Exactly as in [10, Section 4], we may now take an arbitrary linearly compact, commutative ring R with ideal neighborhoods of zero, and show that, since the discrete homomorphs of R decompose uniquely (Proposition 14) these homomorphs decompose concordantly, and therefore that R itself decomposes:

THEOREM 2. *Let R be a linearly compact, commutative ring with ideal neighborhoods of zero. Then R is algebraically and topologically isomorphic to a complete direct sum of a radical ring and primary rings with units, all the summands being linearly compact.*

The converse of Theorem 2 is also true by Proposition 1 and the remark following the definition of linear compactness.

4. Further results and examples. From the definition of linear compactness alone we can verify the following remarks without difficulty:

1. *If M is an R-module linearly compact in one topology, then M is linearly compact in every coarser topology.*

2. *If M is a topological module in two topologies A and B, and if the same submodules are closed in A as in B, then M is linearly compact in topology A if and only if it is linearly compact in topology B.*

3. *A commutative ring R is linearly compact in the discrete topology if and only if it is linearly compact in the R-topology* (where neighborhoods of zero are the finite intersections of nonzero ideals; cf. [11]); *for all ideals are closed in both topologies.*

Remark 3 together with [11, Theorem 1] proves Theorem 3 below; however, we shall give a direct proof, which, together with Proposition 9 will constitute an alternative proof of [11, Theorem 1].

THEOREM 3. *Let R be a complete local ring. Then R is linearly compact in the discrete topology, and hence in every topology in which R is a topological ring.*

Proof. Let R be a complete local ring, M its unique maximal ideal, $F = R/M$ its residue class field. Each of the discrete R-modules R/M, $M/M^2, \cdots, M^i/M^{i+1}, \cdots$ satisfies the ascending chain condition on R-submodules; since these are all actually (unitary) F-modules, they satisfy the ascending chain condition on F-submodules, which proves that they also satisfy the descending chain condition. Thus each of these factor modules is linearly compact, and repeated use of Proposition 9 shows that R/M^i is linearly compact for all i. But R, being complete in its local topology, is the inverse limit of these R-modules R/M^i. Thus R is linearly compact in the local topology by Proposition 4.

Furthermore, every submodule A of R is closed in the local topology; for by [8, Theorem 2] A is the intersection of the neighborhoods $A + M^i$ of A. Remark 2 above then guarantees that R is also linearly compact in the discrete topology.

Theorem 3 shows that the converse of Proposition 5 is false: A complete

local ring is linearly compact in the discrete topology but does not satisfy the minimum condition in general. It also provides a counterexample to the conjecture that every continuous homomorphism of a linearly compact module is open: take the identity mapping of a complete local ring R with the discrete topology onto R with the local topology, both range and domain being considered as modules over the discrete ring R.

These considerations lead to the question: Which commutative rings with units are linearly compact in the discrete topology? By Proposition 14 we may as well restrict ourselves to primary rings. Among such rings are all fields, all complete local rings. I do not know what the full answer is, but we may add one more theorem:[3]

THEOREM 4. *A (generalized) valuation ring is linearly compact in the discrete topology if and only if it is maximal (and if and only if it is linearly compact in its valuation topology, by Remark 2).*

By a maximal valuation ring we mean a (generalized) valuation ring whose quotient field cannot be extended without changing either the value group or the residue class field. By [6, Theorem 4] a valuation ring is maximal if and only if every pseudoconvergent sequence in the quotient field has a limit in the quotient field (for definitions see [6]). It is easily seen that this is equivalent to the assertion that every pseudoconvergent sequence in the ring has a limit in the ring. Therefore it suffices for us to prove that a valuation ring R is linearly compact if and only if every pseudoconvergent sequence in R has a limit in R.

Proof. If: In a valuation ring, the ideals (submodules) are linearly ordered by inclusion. Hence a collection of cosets of ideals (linear varieties) has the finite intersection property if and only if it is also linearly ordered by inclusion. Given such a collection we may extract a well-ordered cofinal subcollection. To show the original collection has a nonvoid intersection it suffices to show this well-ordered subcollection has a nonvoid intersection.

Thus we may begin with a collection $\{F_\rho\}$, indexed by a segment of the ordinal numbers, and monotone: $\sigma > \rho$ implies $F_\sigma \subset F_\rho$ and $F_\sigma \neq F_\rho$. Let A_ρ be the ideal of which F_ρ is a coset. Choose a_ρ in F_ρ but not in $F_{\rho+1}$. Then for $\sigma > \rho$ we have $a_\rho \not\in F_\sigma = a_\sigma + A_\sigma$, $a_\rho - a_\sigma \not\in A_\sigma$, and also $a_\sigma \in F_\rho = a_\rho + A_\rho$, $a_\rho - a_\sigma \in F_\rho$. Hence for $\tau > \sigma > \rho$, $a_\rho - a_\sigma \not\in A_\sigma$ and $a_\sigma - a_\tau \in A_\sigma$, and so $V(a_\sigma - a_\rho) < V(a_\tau - a_\sigma)$ (V is the valuation func-

[3] Professor Kaplansky obtained this result some time ago but did not publish it. I presume his proof was effectively the same as the one given here.

tion), so that the a_ρ form a pseudoconvergent sequence. This sequence has a limit a in R such that $V(a - a_\rho) = V(a_\rho - a_{\rho+1})$. Then $a - a_\rho$ is contained in the ideal generated by $a_\rho - a_{\rho+1}$, which is contained in A_ρ. Thus $a \varepsilon a_\rho + A_\rho = F_\rho$ for every ρ, as desired.

Only if: Let $\{a_\rho\}$ be a pseudoconvergent sequence in R and define $\gamma_\rho = V(a_\rho - a_{\rho+1})$, $A_\rho = \{x \mid V(x) \geqq \gamma_\rho\}$, $F_\rho = a_\rho + A_\rho$. Then the F_ρ form a monotone sequence of linear varieties because $A_\rho \supset A_{\rho+1}$ and $a_\rho \equiv a_{\rho+1} \pmod{A_\rho}$. Thus the F_ρ have a point a in common. Then $V(a - a_\rho) \geqq \gamma_\rho$ for every ρ. If $V(a - a_\rho > \gamma_\rho$, then

$$\gamma_{\rho+1} \leqq V(a - a_{\rho+1}) = \min[V(a - a_\rho),\ V(a_\rho - a_{\rho+1})] = \gamma_\rho,$$

which is impossible. Therefore $V(a - a_\rho) = \gamma_\rho$ and a is a limit of the pseudoconvergent seqeunce $\{a_\rho\}$.

NORTHWESTERN UNIVERSITY.

REFERENCES.

[1] N. Bourbaki, "Les structures fondamentales de l'analyse VI (Algèbre linéaire)," *Actualités Scientifiques et Industrielles*, No. 1032, Paris, 1947.

[2] D. van Dantzig, "Zur topologischen Algebra II. Abstrakte $\mathfrak{b}_\nu$-adische Ringe," *Compositio Mathematica*, vol. 2 (1935), pp. 201-223.

[3] J. Dieudonné, "Linearly compact spaces and double vector spaces over sfields," *American Journal of Mathematics*, vol. 73 (1951), pp. 13-19.

[4] ———, "Matrices semi-finies et espaces localement linéairement compacts," *Journal für die reine und angewandte Mathematik*, vol. 188 (1950), pp. 162-166.

[5] N. Jacobson, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

[6] I. Kaplansky, "Maximal fields with valuations," *Duke Mathematical Journal*, vol. 9 (1942), pp. 303-321.

[7] ———, "Topological rings," *American Journal of Mathematics*, vol. 69 (1947), pp. 153-183.

[8] W. Krull, "Dimensionstheorie in Stellenringen," *Journal für die reine und angewandte Mathematik*, vol. 179 (1938), pp. 204-226.

[9] S. Lefschetz, "Algebraic Topology," *Colloquium Publications*, vol. 27, New York, 1942.

[10] D. Zelinsky, "Rings with ideal nuclei," *Duke Mathematical Journal*, vol. 18 (1951), pp. 431-442.

[11] ———, "Complete fields from local rings," *Proceedings of the National Academy of Sciences*, vol. 37 (1951), pp. 379-381.

SPECIAL TYPES OF LINEAR MAPPINGS OF ALGEBRAS.*

By J. K. Goldhaber.

1. Introduction. Ancochea [2] and Kaplansky [7] considered the problem of semi-automorphisms of simple and semi-simple algebras. They showed (Ancochea for simple algebras over a field of characteristic different from two, Kaplansky for direct sums of simple algebras over arbitrary fields, and Hua [5] for aribtrary sfields) that every semi-automorphism is either an automorphism of an anti-automorphism of each simple component of the algebra. A semi-automorphism (or Jordan automorphism) of an algebra $\mathfrak{A}$ may be defined, in the case that the characteristic of the ground field is different from two, as an additive automorphism, ψ, such that for all $A \varepsilon \mathfrak{A}$, $\psi(A^2) = [\psi(A)]^2$, or what is the same thing, that for all $A, B \varepsilon \mathfrak{A}$, $\psi(AB + BA) = \psi(A)\psi(B) + \psi(B)\psi(A)$.

The problem of this paper may be characterized as follows: to what extent may the conditions on ψ be relaxed and still have the result of the above theorem hold true? We assume throughout that the ground field $\mathfrak{F}$ has characteristic different from two.

In Section **2** we show that if $\mathfrak{A}$ is a total matric algebra then we need only assume $\psi(A^2) = [\psi(A)]^2$ for idempotents and nilpotents of index two. This result, however, does not extend to simple algebras.

In Section **3** we prove the following theorem: Let $\mathfrak{A}$ be a central simple algebra and let ψ be a linear mapping of $\mathfrak{A}$ into $\mathfrak{A}$ such that (i) $\psi(I) = I$ where I is the identity of $\mathfrak{A}$, and (ii) if $f(x)$ is the characteristic function of $A \varepsilon \mathfrak{A}$ then $f[\psi(A)] = 0$, then ψ is either an automorphism or an anti-automorphism of $\mathfrak{A}$. In the proof of this theorem we assume that $\mathfrak{F}$ has at least as many elements as the degree of $\mathfrak{A}$.

The theorem of Section **3** does not hold for semi-simple algebras. However, it is shown in Section **4** that if it is further assumed that the mapping ψ is one to one then the theorem can be extended to semi-simple separable algebras.

I am indebted to the referee for greatly simplifying the proof of Theorem 1.

* Received April 16, 1952; revised July 1, 1952.

2. Linear mappings of total matric algebras which preserve idempotent and nilpotent elements. We shall consider a mapping of a total matric algebra $\mathfrak{M}$ into $\mathfrak{M}$ which satisfies the following conditions:

$$C\dagger: \quad \begin{array}{ll} (1) & \psi \text{ is linear.} \\ (2) & \text{If } M \in \mathfrak{M} \text{ and } M^2 = M \text{ then } [\psi(M)]^2 = \psi(M). \\ (3) & \text{If } M \in \mathfrak{M} \text{ and } M^2 = 0 \text{ then } [\psi(M)]^2 = 0. \end{array}$$

THEOREM 1. *Let $\mathfrak{M}$ be a total matric algebra over a field of characteristic different from two. Let ψ be a mapping of $\mathfrak{M}$ onto $\mathfrak{X} \leqq \mathfrak{M}$ which satisfies conditions $C\dagger$. Then either $\mathfrak{X} = \mathfrak{M}$ and ψ is either an automorphism or an anti-automorphism of $\mathfrak{M}$ or $\mathfrak{X} = 0$ and ψ is the zero mapping.*

Jacobson and Rickart [6] have shown that every Jordan homomorphism of a Jordan simple algebra is either 0 or a Jordan automorphism. Furthermore, every Jordan automorphism of a simple algebra is either an automorphism or an anti-automorphism [2], [7]. Since a total matirc algebra is Jordan simple, Theorem 1 will follow if it is shown that ψ is a Jordan homomorphism.

Since $\mathfrak{M}$ is a total matric algebra there exists a basis for $\mathfrak{M}$ of the form E_{ij} where $E_{ij}E_{km} = \delta_{jk}E_{im}$. Because of the linearity of ψ it will be sufficient to show that $\psi(E_{ij}E_{km} + E_{km}E_{ij}) = \psi(E_{ij})\psi(E_{km}) + \psi(E_{km})\psi(E_{ij})$.

From the fact that ψ satisfies conditions $C\dagger$ and from the following relations

$$\begin{array}{lll}
\text{(i)} & E_{ii}^2 = E_{ii} & \\
\text{(ii)} & E_{ij}^2 = 0 & i \neq j \\
\text{(iii)} & (E_{ij} + E_{km})^2 = 0 & i \neq j,\ j \neq k,\ k \neq m,\ m \neq i \\
\text{(iv)} & (E_{ii} + E_{jj})^2 = (E_{ii} + E_{jj}) & i \neq j \\
\text{(v)} & (E_{ii} + E_{ij})^2 = (E_{ii} + E_{ij}) & i \neq j \\
\text{(vi)} & (E_{jj} + E_{ij})^2 = (E_{jj} + E_{ij}) & i \neq j \\
\text{(vii)} & (E_{ii} - E_{ij} + E_{ji} - E_{jj})^2 = 0 & \\
\text{(viii)} & (E_{ii} + E_{jk} + E_{jj})^2 = (E_{ii} + E_{jk} + E_{jj}) & i \neq j,\ j \neq k,\ k \neq i \\
\text{(ix)} & (E_{ii} + E_{ik} + E_{kj} - E_{ij} + E_{jj})^2 = (E_{ii} + E_{ik} + E_{kj} - E_{ij} + E_{jj}) & \\
 & & i \neq j,\ j \neq k,\ k \neq i
\end{array}$$

one can obtain by simple and direct computations successively

$$\begin{array}{lll}
(1) & \psi(E_{ii})\psi(E_{ii}) = \psi(E_{ii}) & \\
(2) & \psi(E_{ij})\psi(E_{ij}) = 0 & i \neq j
\end{array}$$

(3) $\psi(E_{ij})\psi(E_{km}) + \psi(E_{km})\psi(E_{ij}) = 0 = \psi(E_{ij}E_{km} + E_{km}E_{ij})$
$i \neq j, j \neq k, k \neq m, m \neq i$

(4) $\psi(E_{ii})\psi(E_{jj}) + \psi(E_{jj})\psi(E_{ii}) = 0 = \psi(E_{ii}E_{jj} + E_{jj}E_{ii}) \qquad i \neq j$

(5) $\psi(E_{ii})\psi(E_{ij}) + \psi(E_{ij})\psi(E_{ii}) = \psi(E_{ij}) = \psi(E_{ii}E_{ij} + E_{ij}E_{ii}) \qquad i \neq j$

(6) $\psi(E_{jj})\psi(E_{ij}) + \psi(E_{ij})\psi(E_{jj}) = \psi(E_{ij}) = \psi(E_{jj}E_{ij} + E_{ij}E_{jj}) \qquad i \neq j$

(7) $\psi(E_{ij})\psi(E_{ji}) + \psi(E_{ji})\psi(E_{ij}) = \psi(E_{ii}) + \psi(E_{jj}) = \psi(E_{ij}E_{ji} + E_{ji}E_{ij})$

(8) $\psi(E_{ii})\psi(E_{jk}) + \psi(E_{jk})\psi(E_{ii}) = 0 = \psi(E_{ii}E_{jk} + E_{jk}E_{ii}) \qquad i \neq j, k \neq i$

(9) $\psi(E_{ik})\psi(E_{kj}) + \psi(E_{kj})\psi(E_{ik}) = \psi(E_{ij}) = \psi(E_{ik}E_{kj} + E_{kj}E_{ik}) \qquad i \neq j.$

Equations (1) through (9) may now be used to verify readily that for all basic elements E_{rs}, $\psi(E_{ij}E_{km} + E_{km}E_{ij}) = \psi(E_{ij})\psi(E_{km}) + \psi(E_{km})\psi(E_{ij})$. Hence ψ is a Jordan homomorphism and Theorem 1 is established.

Theorem 1 is not true if the characteristic of the ground field is two. In fact, consider the total matric algebra of order four over $GF(2)$ and let $\psi(\Sigma a_{ij}E_{ij}) = (\Sigma\delta_{ij}a_{ij})(E_{11} + E_{22})$, $a_{ij} \in GF(2)$, $i, j = 1, 2$. ψ is a non-zero mapping which satisfies conditions $C\dagger$ but is neither an automorphism nor an anti-automorphism of the total matric algebra considered.

We also note that Theorem 1 does not extend to simple algebras. For consider the quaternions over the rational field. Any linear mapping which maps zero into zero and the identity into the identity will automatically satisfy conditions $C\dagger$. The difficulty lies in the fact that although ψ may satisfy conditions $C\dagger$ as a mapping of $\mathfrak{A}_{\mathfrak{F}}$ into $\mathfrak{A}_{\mathfrak{F}}$ it will not in general satisfy conditions $C\dagger$ as a mapping of $\mathfrak{A}_{\mathfrak{K}}$ into $\mathfrak{A}_{\mathfrak{K}}$ where $\mathfrak{K}$ is an extension field of $\mathfrak{F}$.

3. Characteristic function preserving linear mappings of central simple algebras. Throughout this and the following section the ground field $\mathfrak{F}$ shall be assumed to have a characteristic different from two and, if the algebra is simple, to contain at least as many distinct elements as the principal degree of the algebra. If the algebra considered is semi-simple then $\mathfrak{F}$ shall be assumed to contain at least as many distinct elements as the maximum of the principal degrees of the simple components of the algebra. We shall be concerned with a mapping ψ of an algebra $\mathfrak{A}$ over $\mathfrak{F}$ into $\mathfrak{A}$ which satisfies the following conditions:

(1) ψ is linear.

$C\dagger'$: (2) $\psi(I) = I$ where I is the identity of $\mathfrak{A}$.

(3) If $f(\lambda)$ is the characteristic function of $A \in \mathfrak{A}$ then $f[\psi(A)] = 0$.

Note that if ψ is a Jordan homomorphism then $\psi(A^n) = [\psi(A)]^n$ for all n. Hence every identity preserving Jordan homomorphism satisfies conditions $C\dagger'$.

THEOREM 2. *Let $\mathfrak{A}$ be a central simple algebra over $\mathfrak{F}$ and let ψ be a mapping of $\mathfrak{A}$ into $\mathfrak{A}$ which satisfies $C\dagger'$. Then ψ is either an automorphism or an anti-automorphism of $\mathfrak{A}$.*

We prove first two lemmas.

LEMMA 1. *Let $\mathfrak{A}$ and ψ be as in Theorem 2. Then* (i) *ψ is a one to one mapping of $\mathfrak{A}$ onto $\mathfrak{A}$ and* (ii) *ψ preserves characteristic functions, i. e. for $A \varepsilon \mathfrak{A}$, A and $\psi(A)$ have the same characteristic function.*

By $C\dagger'_3$ the minimum function of $\psi(A)$ divides the characteristic function, $f(\lambda)$, of A. Now suppose that for some $N \varepsilon \mathfrak{A}$, $N \neq 0$, $\psi(N) = 0$. Since $\mathfrak{A}$ is simple there exists a nilpotent $N' \varepsilon \mathfrak{A}$ such that $N + N'$ is non-singular. Since N' is nilpotent the minimum function of $\psi(N')$ is divisible by λ. Since $N + N'$ is non-singular the minimum function of $\psi(N + N')$ is not divisible by λ. But $\psi(N + N') = \psi(N')$ and we have a contradiction. Hence ψ is one to one and onto $\mathfrak{A}$.

Now let e_i $(i = 1, 2, \cdots, n^2)$ be a basis for $\mathfrak{A}$ over $\mathfrak{F}$. Consider the general element $X = \Sigma x_i e_i$ and let $m(\lambda; x_i)$ be the minimum polynomial of $\mathfrak{A}$ relative to X. Let $h(\lambda; x_i)$ be the minimum polynomial of $\mathfrak{A}$ relative to the general element $\psi(X) = \Sigma x_i \psi(e_i)$. Now $[m(\lambda; x_i)]^n$ is the characteristic function of $\mathfrak{A}$ relative to X, [1], p. 123 or [3], p. 51. Furthermore both $m(\lambda; x_i)$ and $h(\lambda; x_i)$ are monic polynomials in λ. Hence

$$[m(\lambda; x_i)]^n = q(\lambda; x_i) \cdot h(\lambda; x_i) + r(\lambda; x_i)$$

where $r(\lambda; x_i)$ is either 0 or of degree less than n in λ. Since ψ satisfies $C\dagger'$ it is true that $r(\Sigma a_i \psi(e_i); a_i) = 0$ for all choices of $a_i \varepsilon \mathfrak{F}$. Because of the assumption on the number of distinct elements in $\mathfrak{F}$ it now follows that $r(\lambda; x_i) = 0$. Hence $h(\lambda; x_i) \mid [m(\lambda; x_i)]^n$. But since $m(\psi; x_i)$ is a monic irreducible polynomial [1], [3] we have $h(\lambda; x_i) = m(\lambda; x_i)$. Now for $A = \Sigma a_i e_i$, $a_i \varepsilon \mathfrak{F}$, $[m(\lambda; a_i)]^n$ is the characteristic function of A and $[h(\lambda; a_i)]^n$ is the characteristic function of $\psi(A)$. Hence for $A \varepsilon \mathfrak{A}$, A and $\psi(A)$ have the same characteristic function. This proves Lemma 1.

LEMMA 2. *Let $\mathfrak{A}$ and ψ be as in Theorem 2. Let $\mathfrak{K}$ be an extension field of $\mathfrak{F}$ such that $\mathfrak{K}$ splits $\mathfrak{A}$. Extend ψ in the obvious manner to a linear mapping of $\mathfrak{A}_{\mathfrak{K}}$ into $\mathfrak{A}_{\mathfrak{K}}$. Then ψ is a one to one mapping of $\mathfrak{A}_{\mathfrak{K}}$ onto $\mathfrak{A}_{\mathfrak{K}}$*

such that A and $\psi(A)$, considered as elements of $\mathfrak{A}_{\mathfrak{K}}$, have the same characteristic function.

The fact that ψ is one to one follows as in Lemma 1. We have also seen that $m(\lambda; x_i) = h(\lambda; x_i)$. But now for $A = \Sigma a_i e_i$, $a_i \varepsilon \mathfrak{K}$, $m(\lambda; a_i)$ is the characteristic function of A considered as an element of $\mathfrak{A}_{\mathfrak{K}}$ [1], p. 124 or [3], p. 52. Similarly, $h(\lambda; a_i)$ is the characteristic function of $\psi(A)$. Hence A and $\psi(A)$ have the same characteristic function as elements of $\mathfrak{A}_{\mathfrak{K}}$ and Lemma 2 is established.

We now prove Theorem 2 by proving that ψ is either an automorphism or an anti-automorphism of $\mathfrak{A}_{\mathfrak{K}}$. The latter result will follow from Theorem 1 if it is shown that for all $A \varepsilon \mathfrak{A}_{\mathfrak{K}}$, A and $\psi(A)$ have the same minimum function.

Let $X = \Sigma x_i e_i$ be a general element of $\mathfrak{A}_{\mathfrak{K}}$. Then $\psi(X) = Y$ is an $n \times n$ matrix each of its entries being a linear function of the x_i. Now let $(\lambda I - X) = X'$ and let $\psi(X') = \psi(\lambda I - X) = \lambda I - Y = Y'$. Consider $\det(X') - \det(Y')$. From Lemma 2 it follows that for every specialization of the x_i into $\mathfrak{K}$ the latter expression vanishes. From the assumption on the number of distinct elements in the ground field it follows that $\det(X') - \det(Y') \equiv 0$. But then, by a theorem of Frobenius [4] or [8], p. 15, there exist non-singular constant matrices P and Q such that either $PX'Q = Y'$ or $PX'^T Q = Y'$, where X'^T is the transpose of X'. In either case it follows that for every $A \varepsilon \mathfrak{A}_{\mathfrak{K}}$, A and $\psi(A)$ have the same invariant factors, and hence the same minimum function. Hence Theorem 2 is proved.

4. Extension to semi-simple separable algebras. Theorem 2 is not true for semi-simple algebras. For let $\mathfrak{A}$ be the direct sum of two total matric algebras of the same order; $\mathfrak{A} = \mathfrak{M} \oplus \mathfrak{M}$. If $A \varepsilon \mathfrak{A}$ then $A = M_1 \oplus M_2$, $M_1, M_2 \varepsilon \mathfrak{M}$. Define $\psi(A) = M_1 \oplus M_1$. The mapping ψ defined in this manner satisfies conditions $C\dagger'$ but is neither an automorphism nor an anti-automorphism of $\mathfrak{A}$. We have however

THEOREM 3. *Let $\mathfrak{A}$ be a semi-simple separable algebra and let ψ be a one to one mapping of $\mathfrak{A}$ onto $\mathfrak{A}$ such that ψ satisfies conditions $C\dagger'$. Then ψ induces on each simple component of $\mathfrak{A}$ either an automorphism or an anti-automorphism.*

Since ψ is one to one onto $\mathfrak{A}$ it follows as in the second part of Lemma 1 that ψ preserves characteristic functions. We now show that it will be sufficient to prove the theorem for the case that $\mathfrak{A}$ is a direct sum of total matric

algebras. If the ground $\mathfrak{F}$ is finite then $\mathfrak{A}$ is a direct sum of total matric algebras, [1]. Suppose then that $\mathfrak{F}$ is infinite. $\mathfrak{A}$ is a direct sum of simple algebras $\mathfrak{A}_i$. Let e_{ij} $(j=1,2,\cdots,n_i^2)$ be a basis for $\mathfrak{A}_i$ over $\mathfrak{F}$. Let $f(\lambda;x_{ij})$ be the characteristic function of $\mathfrak{A}$ relative to the general element $X=\Sigma x_{ij}e_{ij}$ and let $g(\lambda;x_{ij})$ be the characteristic function of $\mathfrak{A}$ relative to $\psi(X)$. Since ψ preserves characteristic functions and since $\mathfrak{F}$ is infinite, $f(\lambda;x_{ij})\equiv g(\lambda;x_{ij})$. But now $f(\lambda;x_{ij})=\prod_i m_i(\lambda;x_{ij})^{n_i}$, where $m_i(\lambda;x_{ij})$, of degree n_i in λ, is the minimum function of $\mathfrak{A}_i$ and $m_i(\lambda;x_{ij})$ contains no x_{hj}, $h\neq i$. The minimum function of $\mathfrak{A}$ is $\prod_i m_i(\lambda;x_{ij})$. Because of the identity of f and g it follows that the minimum function of $\mathfrak{A}$ relative to X is identical with the minimum function of $\mathfrak{A}$ relative to $\psi(X)$. Hence, as above, if $\mathfrak{K}$ is a splitting field for $\mathfrak{A}$, ψ will be a one to one mapping which satisfies conditions $C\dagger'$ as a mapping of $\mathfrak{A}_{\mathfrak{K}}$ onto $\mathfrak{A}_{\mathfrak{K}}$. Clearly if the theorem is true for $\mathfrak{A}_{\mathfrak{K}}$ it is also true for $\mathfrak{A}$.

Suppose then that $\mathfrak{K}$ is the algebraic closure of $\mathfrak{F}$ and that

$$\mathfrak{A}_{\mathfrak{K}}=\mathfrak{M}_1\oplus\mathfrak{M}_2\oplus\cdots\oplus\mathfrak{M}_k.$$

Theorem 3 will follow from Theorem 2 if it is shown that $\psi(\mathfrak{M}_h)\leqq\mathfrak{M}_h$ and $\psi(I_h)=I_h$ where I_h is the identity of $\mathfrak{M}_h$ for $h=1,2,\cdots,k$.

Let I_h be the identity of $\mathfrak{M}_h$. We show first that if $\psi(I_h)\,\varepsilon\,\mathfrak{M}_h$ then $\psi(M)\,\varepsilon\,\mathfrak{M}_h$ for all $M\,\varepsilon\,\mathfrak{M}_h$. For consider $\psi(M)=M_1+M_2+\cdots+M_h+\cdots+M_k$ where $M_i\,\varepsilon\,\mathfrak{M}_i$ and $M\,\varepsilon\,\mathfrak{M}_h$. Suppose now that (say) $M_1\neq 0$. Then there exists a nilpotent matrix $N_1\,\varepsilon\,\mathfrak{M}_1$ such that M_1+N_1 is non-singular in $\mathfrak{M}_1$. Since ψ is one to one onto $\mathfrak{A}$ there exists an $A\,\varepsilon\,\mathfrak{A}$, $A=A_1+A_2+\cdots+A_h+\cdots+A_k$ where $A_i\,\varepsilon\,\mathfrak{M}_i$ and A_i is nilpotent, such that $\psi(A)=N_1$. Now choose $a\,\varepsilon\,\mathfrak{K}$ such that both M_h+aI_h and $M+A_h+aI_h$ are non-singular in $\mathfrak{M}_h$. Clearly $M+A+aI_h$ has exactly n_h non-zero characteristic roots, where n_h^2 is the order of $\mathfrak{M}_h$. But

$$\begin{aligned}&\psi(M+A+aI_h)\\&\qquad=[M_1+N_1]+M_2+\cdots+[M_h+aI_h]+M_{h+1}+\cdots+M_k\end{aligned}$$

has at least n_h+1 non-zero characteristic roots. Since ψ preserves characteristic functions, we have a contradiction. Hence for $i\neq h$, $M_i=0$, and $\psi(M)\,\varepsilon\,\mathfrak{M}_h$ for $M\,\varepsilon\,\mathfrak{M}_h$.

We now show that $\psi(I_h)\,\varepsilon\,\mathfrak{M}_h$. Suppose that $\psi(I_h)=\sum_{i\,\varepsilon\,S}J_i$ where $J_i\neq 0$ and $J_i\,\varepsilon\,\mathfrak{M}_i$ and S is a subset of the integers $1,2,\cdots,k$. Let n_i be the degree of $\mathfrak{M}_i$. Since ψ preserves characteristic functions it follows that $\sum_{i\,\varepsilon\,S}n_i\geqq n_h$. Now if $\sum_{i\,\varepsilon\,S}n_i>n_h$ then there exist nilpotent $N_i\,\varepsilon\,\mathfrak{M}_i$ such that

$\sum_{i \varepsilon S} (J_i + N_i)$ has exactly $\sum_{i \varepsilon S} n_i$ non-zero characteristic roots. But $\sum_{i \varepsilon S} (J_i + N_i)$ is the map of an element which has at most n_h non-zero roots, which is impossible. Hence $\sum_{i \varepsilon S} n_i = n_h$. Furthermore, we may show, as above, that for all $M \varepsilon \mathfrak{M}_h$, $\psi(M)$ is an element of the direct sum of $\mathfrak{M}_i$, $i \varepsilon S$. Since ψ preserves linear independence we must have $\sum_{i \varepsilon S} n_i^2 \geqq n_h^2$. In view of the previously established equality this is only possible if the set S consists of only one integer r and $n_r = n_h$. We may therefore consider $\psi(I_h)$ as mapping into $\mathfrak{M}_h$. Since ψ maps the identity of $\mathfrak{A}$ into itself, it must, in view of the above discussions, map I_h into I_h. Hence ψ maps each $\mathfrak{M}_i$ into $\mathfrak{M}_i$ and ψ satisfies conditions $C\dagger'$ on each $\mathfrak{M}_i$. Our theorem now follows from Theorem 2.

Theorem 3 remains valid if we replace the assumption that ψ is one to one by the assumption that ψ preserves characteristic functions. We note also that if we assume that X and $\psi(X)$ have the same characteristic function, where X is a general element, then the requirement on the number of distinct elements in $\mathfrak{F}$ becomes unnecessary.

UNIVERSITY OF CONNECTICUT.

REFERENCES.

[1] A. A. Albert, *Structure of Algebras*, New York, 1939.

[2] G. Ancochea, "On semi-automorphisms of division algebras," *Annals of Mathematics*, vol. 48 (1947), pp. 147-153.

[3] M. Deuring, "Algebren," *Ergebnisse der Mathematische*, vol. 4 (1935).

[4] G. Frobenius, *Sitzungsberichte Akademie der Wissenschaft* (1897), pp. 994-1015.

[5] L. K. Hua, "On the automorphisms of a sfield," *Proceedings of the National Academy of Science*, vol. 35 (1949), pp. 386-389.

[6] N. Jacobson and C. E. Rickart, "Jordan homomorphisms of rings," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 479-502.

[7] I. Kaplansky, "Semi-automorphism of rings," *Duke Mathematical Journal*, vol. 14 (1947), pp. 521-525.

[8] C. C. MacDuffee, "Theory of matrices," *Ergebnisse der Mathematische*, vol. 2 (1933).

ON SIMPLE ALTERNATIVE RINGS.*

By Erwin Kleinfeld.

1. Introduction. Since alternative rings were first defined by Max Zorn [9] in 1930, this subject has made great strides. In fact the point has been reached where for a wide class of rings we can say much more by assuming they are alternative and not associative than by assuming the associative law. Such a class must of necessity be limited by the fact that the direct sum of an associative and an alternative ring is again an alternative ring.

Recent investigations by L. A. Skornyakov, R. H. Bruck and the author have shown that the *Cayley-Dickson division algebras are the only alternative, not associative division rings* [3], [5], [7]. Assuming characteristic not 2 the author has extended this result to simple alternative rings without nilpotent elements [4]. On the other hand A. A. Albert has shown that *all simple, alternative, not associative rings, having a non-trivial idempotent, are Cayley-Dickson vector matrix algebras* [2]. This leads one to suspect that perhaps there exist no other simple, alternative, not associative rings. For finite dimensional algebras the conjecture can be answered in the affirmative, wtih the help of R. D. Schafer's work [6]. The present paper throws some new light on this question, but does not settle it entirely.

Throughout this paper, *R will denote an alternative, not associative, simple ring of characteristic not* 2. From [2] we are able to deduce that *R is a Cayley-Dickson algebra (this includes both the vector matrix and division algebras) if and only if it contains an element x not in its center, such that x^2 is a non-zero element in the center.* [3] is filled with identities which make it plausible that such elements should exist. In particular, even powers of associators and commutators have certain desirable properties which enable us to satisfy this condition under suitable assumptions.

We are able to show that *R is a Cayley-Dickson division algebra if and only if $(a, b) = 0$ implies $(a, b, R) = 0$.* This condition is more satisfactory than the assumption of no nilpotent elements, since it is weaker in the case of algebras and there exist numerous algebras with radical satisfying it. A motivation for this assumption is to be found in [3], in particular Lemmas 3. 3 and 4. 1.

* Received April 7, 1952.

A necessary and sufficient condition that R be a Cayley-Dickson algebra is the existence of elements a, b, c in R which are pairwise anti-commutative and whose associator (a, b, c) is not a divisor of zero. Identity 2.20 of [3] shows how to obtain an abundance of anti-commutative elements. We discover incidentally that *for any alternative ring S containing pairwise anti-commutative elements x, y, z the identity $(S, x^2)(x, y, z) = 0$ holds.* This leads to a substantial simplification of the division ring proof for characteristic not 2, by showing directly that squares of commutators are in the center of the ring.

2. Preliminaries. We begin with the application of Albert's result. Kaplansky pointed out to the author in conversation that it is sufficient to show the existence of a field in R, quadratic over the center C of R, in order to establish the finite dimensionality of R. For let F be such a field. We form the Kronecker product $K = R \times F$ over C. K is then a simple, alternative, not associative ring containing a non-trivial idempotent, hence a Cayley-Dickson vector matrix algebra over its center C, by Albert's result. Consequently R, a simple, non-associative, alternative subring of K, must have been a Cayley-Dickson algebra to begin with.

THEOREM 2.1. *A necessary and sufficient condition that R be a Cayley-Dickson algebra over its center C is that there exist an element x in R but not in C, such that x^2 is a non-zero element in C.*

Proof. Such an element can always be found in a Cayley-Dickson algebra. Conversely, given x, since $x^2 = c$ is in C, C is a field. We form $F = C + Cx$. If the equation $x^2 - c = 0$ is irreducible, then F is a quadratic field over C and the previous argument applies. If, on the other hand, $x^2 - c = 0$ is reducible, then there exists an element d in C such that $(x - d)(x + d) = 0$. But then $(x + d)^2 = 2d(x + d)$ and we obtain that $y = (x + d)/2d$ has the property $y^2 = y$. Since x itself is not in C, y is a non-trivial idempotent. In either case then R must be a Cayley-Dickson algebra over C.

3. A new identity. We begin by reviewing a number of results from [3]. Let S be any alternative ring and w, x, y, z arbitrary elements of S. Then $f(w, x, y, z)$ is defined by the identity

$$(1) \qquad (wx, y, z) = f(w, x, y, z) + x(w, y, z) + (x, y, z)w.$$

It can be shown that

(2) $f(w, x, y, z)$ *is skew-symmetric in its four variables.*

Moreover the following identities are satisfied:

$$f(w,x,y,z)=((w,x),y,z)+(w,x,(y,z)), \tag{3}$$

$$3f(w,x,y,z)=(w,(x,y,z))-(x,(y,z,w)) + (y,(z,w,x))-(z,(w,x,y)). \tag{4}$$

From the definition of the commutator we obtain

$$(xy,z)=x(y,z)+(x,z)y+3(x,y,z). \tag{5}$$

Also by some calculations involving $f(w,x,y,z)$ it can be shown that

$$((x,y,z),x,y)=(x,y)(x,y,z)=-(x,y,z)(x,y). \tag{6}$$

THEOREM 3.1. *Let a, b, c be elements in S such that*

$$(c^2,a,S)=(c^2,b,S)=(c^2,a)=(c^2,b)=0.$$

Then $(S,c^2)(a,b,c)=0$.

Proof. Let p, q be any elements of S. Then

$$(c^2,(a,b,c),p)=(c,c(a,b,c)+(a,b,c)c,p)=(c,(a,b,c^2),p)=0.$$

By means of (3), with $w=c^2$, $x=a$, $y=b$, $z=q$ we obtain

$$f(c^2,a,b,q)=((c^2,a),b,q)+(c^2,a,(b,q))=0.$$

By means of (4), with $w=c^2$, $x=a$, $y=b$, $z=q$, therefore,

$$\begin{aligned} 0&=3f(c^2,a,b,q)\\ &=(c^2,(a,b,q))-(a,(b,q,c^2))+((b,q,c^2,a))-(q,(c^2,a,b))\\ &=(c^2,(a,b,q)). \end{aligned}$$

Now we let $q=p$, $q=pc$ and $q=(p,c)$ in succession and obtain

(7) $(c^2,(a,b,p))=0$, (8) $(c^2,(a,b,pc))=0$,

(9) $(c^2,(a,b,(p,c)))=0$.

If we let $x=a$, $y=b$, $z=c^2$ in (5), then we get

$$(ab,c^2)=a(b,c^2)+(a,c^2)b+3(a,b,c^2)=0.$$

Similarly $(ba,c^2)=0$, and therefore

$$((a,b),c^2)=0. \tag{10}$$

Consequently

$$c^2((a,b),c,p)=((a,b)c^2,c,p)=(c^2(a,b),c,p)=((a,b),c,p)c^2,$$

making use of (10). From this follows

$$(11) \qquad (c^2, ((a, b), c, p)) = 0.$$

Let $w = a$, $x = b$, $y = c$, $z = p$ in (3) to obtain

$$f(a, b, c, p) = ((a, b), c, p) + (a, b, (c, p)).$$

Now by means of (11) and (9), c^2 commutes with the right hand side of the last equation. Therefore

$$(12) \qquad (c^2, f(a, b, c, p)) = 0.$$

On the other hand (1), with $w = c$, $x = p$, $y = a$, $z = b$ implies that

$$(13) \qquad (cp, a, b) = f(c, p, a, b) + p(c, a, b) + (p, a, b)c.$$

Now put $x = (p, a, b)$, $y = c$, $z = c^2$ in (5) to obtain

$$((p, a, b)c, c^2) = (p, a, b)(c, c^2) + ((p, a, b), c^2)c + 3((p, a, b), c, c^2).$$

Therefore by means of (7) it follows that

$$(14) \qquad ((p, a, b)c, c^2) = 0.$$

Combining (8), (12), (2) and (14), and commuting each side of (13) with c^2 yields

$$(15) \qquad (p(a, b, c), c^2) = 0.$$

If we put $x = p$, $y = (a, b, c)$, $z = c^2$ in (5) we obtain

$$0 = (p(a, b, c), c^2) = p((a, b, c), c^2) + (p, c^2)(a, b, c) + 3(p, (a, b, c), c^2)$$
$$= (p, c^2)(a, b, c).$$

Since p was taken to be an arbitrary element in S, $(S, c^2)(a, b, c) = 0$ and the proof of the theorem is completed.

4. $(a, b) = 0$ implies $(a, b, R) = 0$. Throughout this section we assume in addition that whenever two elements a, b in R commute then $(a, b, R) = 0$. This condition automatically rules out the Cayley-Dickson vector matrix algebras.

In order to prove the finite dimensionality of R over its center C we require some results of [3] and [4]. *Let $u = (x, y, z)$, where x, y, z are arbitrary elements in R. Then*

(16) $(u^2, ux) = (u^2, xu) = 0$,

(17) $(a, b, R) = 0$ *implies* $(a, b) = 0$,

(18) *the nucleus N of R coincides with C and is either zero or a field.*

With the aid of (17) and the additional assumption on R it can be shown that

(19) *for an arbitrary element q in R the set T of all t in R such that $t(q,R,R)=(q,R,R)t=(q,tR,R)=(q,Rt,R)=(q,t,R)=0$, is a two-sided ideal of R.*

From (19) and Lemma 3.3 of [3] it follows that

(20) *if $(p,x)=(p,y)=0$ and $(x,y)\neq 0$, then p is in C.*

We first prove a couple of Lemmas:

LEMMA 4.1. *If r,s are elements of R such that $r^2=s^2=rs+sr=0$, then $rs=0$.*

Proof. Suppose $rs\neq 0$. Then $(r,s)=2rs\neq 0$. But $(rs,r)=(rs,s)=0$. We use (20) to obtain that rs is in C. But $(rs)^2=0$. Because of (18) we have obtained a contradiction. This proves the Lemma.

LEMMA 4.2. *Let x,y,z be arbitrary elements in R and $u=(x,y,z)$. Then u^2 is an element of C.*

Proof. Suppose u^2 is not in C. Clearly $(u^2,u)=0$. By means of (16) we obtain $(u^2,ux)=(u^2,xu)=0$. Consequently we use (20) to show that $(u,ux)=(u,xu)=0$, from which $(u^2,x)=0$ follows. By a similar argument $(u^2,y)=0$. We have $u\neq 0$, so that $(x,y)\neq 0$ by hypothesis. If we use (20) again, we obtain that u^2 is in C. This is a contradiction, and so u^2 is in C anyway.

We now establish the main result of this section:

THEOREM 4.3. *R is a Cayley-Dickson division algebra over C.*

Proof. If R contains no nilpotent elements, we are done, because of the main result of [4]. Otherwise we can find an element in R such that $x\neq 0$, $x^2=0$. Since C is either zero or a field, x cannot be in C.

Suppose that for every r,s in R, $(x,r,s)^2=0$. Since

$$0=(x^2,r,s)=x(x,r,s)+(x,r,s)x,$$

we can use Lemma 4.1 to obtain $x(x,r,s)=(x,r,s)x=0$. If we put $q=x$ in (19), then it is easily verified that x is in T, so that $T=R$. But then $(x,R,R)=0$ and this means x is an element of N. However x is not in C

and this contradicts (18). Therefore there must exist r_1, s_1 in R with the property $(x, r_1, s_1)^2 \neq 0$. Let $v = (x, r_1, s_1)$. Then by Lemma 4.2 v^2 is in C and C is a field. Suppose that v is also in C. Since $v \neq 0$, $(x, r_1) \neq 0$ by hypothesis. However putting $x = x$, $y = r_1$, $z = s_1$ in (6), we obtain $0 = (v, x, r_1) = (x, r_1)v$. Since C is a field this implies $(x, r_1) = 0$, a contradiction. Thus v is not an element of C and we may now use Theorem 2.1 to show that R is a Cayley-Dickson algebra. Since the vector matrix algebra has already been ruled out, R must be a Cayley-Dickson division algebra. This completes the proof.

It is of interest to note that the additional hypothesis on R in this section could have been replaced by either (20) or Lemma 4.1 and Theorem 4.3 would still be true. Since the arguments involved are very similar to the preceding ones, we omit the details.

5. Further results. We indicate first how Theorem 3.1 leads to a simplification of the division ring proof. Let S be a non-associative, alternative division ring of characteristic not 2. We can choose x, y in S such that $v = (x, y) \neq 0$. Because of (17) there exists an element z in S such that $u = (x, y, z) \neq 0$, and moreover (6) tells us that $uv + vu = 0$. Consequently $(u, v) \neq 0$ and we can use (17) again to obtain the existence of an element t in S such that $w = (u, v, t) \neq 0$. Also $uw + wu = vw + wv = 0$ and $(u, v, w) = 2uv \cdot w \neq 0$. We apply Theorem 3.1, with $a = w$, $b = u$, $c = v$, and obtain that v^2 is in the center C of S. In fact the square of any commutator must be in C. At this point we can apply any one of a number of results to prove S is a Cayley-Dickson division algebra. In particular [1], [8], or Theorem 2.1 of the present paper could be used. Of these [8] seems most appropriate in this situation.

Finally we sketch the proof of the following result:

THEOREM 5.1. *A necessary and sufficient condition that R be a Cayley-Dickson algebra is that there exist pairwise anti-commutative elements a, b, c in R such that $u = (a, b, c)$ is not a divisor of zero.*

Proof. Since u, a, b satisfy the hypothesis of Theorem 3.1 we obtain $(u^2, R)(u, a, b) = 0$. By hypothesis it follows that $2ab \cdot u$ annihilates (u^2, R). By some straightforward calculations it follows that ab annihilates (u^2, R), and therefore so do bc and ac. We show in successive steps that $((u^2, R), a, b, c) = 0$, so that $a \cdot bc$ and $ab \cdot c$ annihilate (u^2, R). We conclude that $(u^2, R)u = 0$. Since u is not a divisor of zero, $(u^2, R) = 0$. Even when R has

characteristic 3 it follows that u^2 is in C. By hypothesis $ua + au = 0$, so that u cannot be in C. We then apply Theorem 2.1 and the proof is complete.

Added in proof (December 11, 1952). The conjecture about simple alternative rings made in the introduction has recently been proved by the author and the proof now awaits publication.

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

[1] A. A. Albert, "Absolute-valued algebraic algebras," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 763-768.

[2] ———, "On simple alternative rings," *Canadian Journal of Mathematics*, vol. 4 (1952), pp. 129-135.

[3] R. H. Bruck and Erwin Kleinfeld, "The structure of alternative division rings," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 878-890.

[4] Erwin Kleinfeld, "An extension of the theorem on alternative division rings," *ibid.*, vol. 3 (1952), pp. 348-351.

[5] ———, "Alternative division rings of characteristic 2," *Proceedings of the National Academy of Sciences*, vol. 37 (1951), pp. 818-820.

[6] R. D. Schafer, "Alternative algebras over an arbitrary field," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 549-555.

[7] L. A. Skornyakov, "Alternative fields," *Ukrain. Mat. Zurnal*, vol. 2 (1950), pp. 70-85 (Russian).

[8] M. F. Smiley, "A remark on a theorem of Marshall Hall," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 342-343.

[9] Max Zorn, "Theorie der alternativen Ringe," *Abhandlungen des Mathematischen Seminars der Hamburgischen Universität*, vol. 8 (1930), pp. 123-147.

A GENERALIZATION OF A THEOREM OF JACOBSON III.*

By I. N. HERSTEIN.

A well known theorem of Jacobson ([3], Theorem 11) states that if in a ring R every element satisfies an equation of the form $x^{n(x)} = x$ with $n(x) > 1$ depending on x, then R is commutative. In a recent paper ([1]) we proved the following generalization of this: "Let R be a ring with center Z, and suppose that for every $x \in R$, $x^n - x \in Z$ where n is a *fixed* integer larger than 1; then R is commutative." In [2] we extended this result to the case that n is no longer fixed, depending on x for its values but where n is uniformly bounded for all x. We have succeeded here in removing all such restrictions and now prove the

THEOREM. *Let R be a ring with center Z. Suppose that every $x \in R$ satisfies $x^{n(x)} - x \in Z$, where $n(x) > 1$ is an integer which depends on x. Then R is necessarily commutative.*

In the remainder of this paper R will always denote a ring with center Z such that $x^{n(x)} - x \in Z$ for every $x \in R$, with $n(x) > 1$ an integer depending on x. Whenever we use the word ideal in an unqualified fashion we will always mean a two-sided ideal.

1. The division and semi-simple ring cases. To establish the result for the division ring case we make use of the following recent theorem due to Krasner [5].

THEOREM 1. (Krasner) *Let K be a finite extension of a field k. If every $x \in K$ satisfies an equation of the form $x^{n(x)} - x \in k$, where $n(x) > 1$ is an integer depending on x, then k is an algebraic extension of a finite field.*

Now suppose that R is a division ring. For $x \in R$ consider the field $Z(x)$ obtained by adjoining x to the center Z. Since every $t \in Z(x)$ satisfies an equation of the form $t^{n(t)} - t \in Z$, and since $Z(x)$ is a finite extension of Z, Krasner's theorem applies. But then Z is algebraic over a finite field, so that every $x \in R$ is algebraic over this finite field. Thus every $x \in R$ satisfies

* Received April 22, 1952.

$x^{t(x)} = x$ for some $t(x) > 1$. We are then thrown back to a situation where Jacobson's theorem can act and we have

THEOREM 2. *If R is a division ring, then R is commutative.*

We are now in a position to start the step-by-step climb to establish the theorem for semi-simple rings using as our ladder the Jacobson structure theorems [4]. In order to do so we need a few preliminary lemmas. In these lemmas we make no assumption of semi-simplicity for R.

LEMMA 3. *For each $x \varepsilon R$ there exist arbitrarily large $n(x)$ with $x^{n(x)} - x \varepsilon Z$.*

For if $x^n - x \varepsilon Z$, $n > 1$, and $(x^n)^m - x^n \varepsilon Z$, $m > 1$, then $x^{nm} - x \varepsilon Z$. Continuing in this way we have the lemma.

LEMMA 4. *If $x \varepsilon R$ is nilpotent, then $x \varepsilon Z$.*

For suppose that $x^t = 0$. By Lemma 3 we can choose an $n > t$ so that $x^n - x \varepsilon Z$. Since $x^n = 0$, x must be in Z.

Suppose on the other hand that $e \varepsilon R$ and $e^2 = e$. For any $x \varepsilon R$, $(xe - exe)^2 = (ex - exe)^2 = 0$. This puts $xe - exe$ in Z. Thus

$$0 = e(xe - exe) = (xe - exe)e = xe - exe;$$

hence $xe = exe$. Similarly $ex = exe$, from which we deduce that $ex = xe$, and we thus have

LEMMA 5. *If $e \varepsilon R$ is an idempotent, then $e \varepsilon Z$.*

Consider $b \varepsilon R$, where b is a divisor of zero in R, that is, $bx = 0$ with $b \neq 0$, $x \neq 0$. Then $b(x^{n(x)} - x) = 0$. If $x^{n(x)} - x \neq 0$, b annihilates a non-zero center element. If, on the contrary, $x^{n(x)} = x$, then as is easily verified $e = x^{n(x)-1} \neq 0$ is an idempotent, and therefore must be in Z. We also have $be = 0$. Thus b again annihilates a non-zero center element. Summarizing, we can say

LEMMA 6. *If $b \varepsilon R$ is a divisor of zero in R, then b annihilates a non-zero center element of R.*

Let us now turn to the case that R is a primitive ring (cf. [4]). Then R has a maximal right ideal ρ which contains no non-zero ideal of R. Let $x \neq 0 \varepsilon \rho$. Since $x^{n(x)} - x \varepsilon Z$, $(x^{n(x)} - x)R$ is an ideal of R and is contained in ρ. The result of this is that $(x^{n(x)} - x)R = (0)$; but the primitivity of R allows this only if $x^{n(x)} = x$. This leads to the conclusion that $e = x^{n(x)-1}$

is an idempotent, and therefore in Z by Lemma 5; thus $eR \subset \rho$ is an ideal of R contained in ρ, and since $e^2 = e \neq 0 \varepsilon eR$, $eR \neq (0)$. This contradiction can be avoided only if $\rho = (0)$. But then R is a division ring, in which case Theorem 2 forces it to be commutative. So we have proved

THEOREM 7. *If R is primitive, then it is also commutative.*

Any ring R, semi-simple in the sense of Jacobson, is isomorphic to a subdirect sum of primitive rings R_i each of which is a homomorphic image of R. The property of R that $x^{n(x)} - x \varepsilon Z$ is thus inherited by each of the R_i. These, in turn, must be commutative, by Theorem 7, since they are primitive. R, being a subdirect sum of commutative rings is itself commutative. Hence

THEOREM 8. *If R is semi-simple then it must be commutative.*

Suppose, then, that R has N as its Jacobson radical. R/N is semi-simple, so it is also commutative. All the commutators are thus located in N. In other words

THEOREM 9. *For all $x, y \varepsilon R$, $xy - yx \varepsilon N$.*

2. The case R is subdirectly irreducible. A ring R is said to be subdirectly irreducible if the intersection of all its non-zero ideals is not merely the zero ideal. Every ring is isomorphic to a subdirect sum of subdirectly irreducible rings. In order to prove our theorem for general rings, in this way, it is sufficient to prove it for subdirectly irreducible rings. *Henceforth in this paper R will always denote a subdirectly irreducible ring with center Z where $x^{n(x)} - x \varepsilon Z$ for all $x \varepsilon R$.*

Let S be the intersection of the non-zero ideals of R. Then $S \neq (0)$.

Consider $s \neq 0 \varepsilon S$. If we suppose R to be non-commutative, then by Theorem 8, $N \neq (0)$, so that $S \subset N$. Since $s \varepsilon N$, $s^{n(s)} \neq s$ (cf. [4]), which implies that $s^{n(s)} - s \varepsilon Z \cap S$. $T = S(s^{n(s)} - s)$ is thus an ideal of R, and since $T \subset S$, we either have $T = (0)$ or $T = S$. Since $S \subset N$ and $s^{n(s)} - s \varepsilon S$, the possibility that $T = S$ is ruled out ([4]). Consequently $T = (0)$. From this $s(s^{n(s)} - s) = 0$, which leads to $s^2 = s^{n(s)+1} = s^{2n(s)}$; but this again is only possible if $s^2 = 0$, since $s \varepsilon N$. So Lemma 4 tells us that $s \varepsilon Z$. For all $s \varepsilon S$, since Ss is now an ideal of R and $Ss \neq S$, we are left with $Ss = (0)$; that is $S^2 = (0)$. Altogether we have shown

THEOREM 10. *If R is subdirectly irreducible and non-commutative, then $S \subset Z$ and $S^2 = (0)$.*

We can now prove the key

THEOREM 11. *In R, $xy - yx$ is nilpotent for every $x, y \varepsilon R$.*

Proof. Suppose $c = uv - vu \neq 0$ is not nilpotent. Since $S^2 = (0)$, $c^i \not\varepsilon S$ for any integer i. Using Zorn's lemma we can find an ideal $U \neq (0)$ of R so that

(1) $c^i \not\varepsilon U$ for any integer i

(2) if W is an ideal of R and $W \supset U$, $W \neq U$, then $c^k \varepsilon W$ for some integer k.

Let $\bar{R} = R/U$. In $\bar{R}$ we have

(a) $\bar{x}^{n(\bar{x})} - \bar{x} \varepsilon \bar{Z}$, the center of $\bar{R}$, for all $\bar{x} \varepsilon \bar{R}$,

(b) if $\bar{V} \neq (0)$ is an ideal of $\bar{R}$ then $0 \neq \bar{c}^{k(\bar{V})} \varepsilon \bar{V}$ for an appropriate integer $k(\bar{V})$.

We claim that $\bar{R}$ has no non-zero divisors of zero. For if $\bar{a}\bar{b} = 0$, $\bar{a} \neq 0$, $\bar{b} \neq 0$, by Lemma 6 we may assume that $\bar{b} \varepsilon \bar{Z}$. Let $T(\bar{b}) = \{\bar{x} \varepsilon \bar{R} \mid \bar{x}\bar{b} = 0\}$. $T(\bar{b})$ is an ideal of $\bar{R}$, since $\bar{b} \varepsilon \bar{Z}$, and since it contains $\bar{a} \neq 0$, $T(\bar{b}) \neq (0)$. Hence $\bar{c}^k \varepsilon T(\bar{b})$ for some suitable integer k. Let $\bar{d} = \bar{c}^{kn} - \bar{c}^k \varepsilon \bar{Z}$, where $n > 1$. Since $\bar{c} \varepsilon \bar{N}$ (Theorem 9) $\bar{d} \neq 0$. Also $\bar{d} \varepsilon T(\bar{b})$. Consider $T(\bar{d}) = \{\bar{x} \varepsilon \bar{R} \mid \bar{d}\bar{x} = 0\}$. Since $\bar{d} \varepsilon \bar{Z}$, $T(\bar{d})$ is an ideal of $\bar{R}$; moreover it contains $\bar{b}$ so it is not just the ideal (0). As a consequence $\bar{c}^j \varepsilon T(\bar{d})$ for some integer j. This leads to $0 = \bar{c}^j\bar{d} = \bar{c}^{kn+j} - \bar{c}^{k+j}$, which is impossible since $\bar{c} \varepsilon N$ and is not nilpotent. So we are forced to assume that $\bar{R}$ has no non-zero divisors of zero.

Let $\bar{V} \neq (0)$ be an ideal of $\bar{R}$. Then $\bar{c}^i \varepsilon \bar{V}$ for some integer i. By Lemma 3 we may suppose that $\bar{c}^{n(\bar{c})} - \bar{c} \varepsilon Z$, with $n(\bar{c}) > i$. Thus if we let $n(\bar{c}) = n$, $\bar{c}\bar{x} - \bar{x}\bar{c} = \bar{c}^n\bar{x} - \bar{x}\bar{c}^n \varepsilon \bar{V}$. So $\bar{c}\bar{x} - \bar{x}\bar{c} \varepsilon \bar{V}$ for all ideals $\bar{V} \neq (0)$ of $\bar{R}$. $\bar{R}$ can not be subdirectly irreducible, for if it were, since it is also not commutative $(\bar{c} = \bar{u}\bar{v} - \bar{v}\bar{u} \neq 0)$ by Theorem 10 its minimal ideal would be nilpotent, which possibility is ruled out by the absence of zero divisors in $\bar{R}$. So we are left with only one alternative, namely that $\bar{c}\bar{x} - \bar{x}\bar{c} = 0$ for all $\bar{x} \varepsilon \bar{R}$; that is $\bar{c} \varepsilon \bar{Z}$. For any $\bar{x} \neq 0$ in $\bar{R}$ let $\bar{r} = \bar{c}\bar{x}$. Since $\bar{c} \varepsilon \bar{Z}$, $\bar{r}^i = \bar{c}^i\bar{x}^i$ for all integers i. So for every ideal $\bar{V} \neq (0)$ of $\bar{R}$, $\bar{r}^i = \bar{c}^i\bar{x}^i \varepsilon \bar{V}$ for appropriate i. The argument used for proving $\bar{c} \varepsilon \bar{Z}$ then applies equally well to $\bar{r}$, so we have $\bar{r} \varepsilon \bar{Z}$. So for all $\bar{y} \varepsilon \bar{R}$, $\bar{r}\bar{y} = (\bar{c}\bar{x})\bar{y} = \bar{y}(\bar{c}\bar{x}) = \bar{c}\bar{y}\bar{x}$, from which it follows that $\bar{c}(\bar{x}\bar{y} - \bar{y}\bar{x}) = 0$. Since $\bar{R}$ is free of divisors of zero we must have that $\bar{x}\bar{y} - \bar{y}\bar{x} = 0$ for all $\bar{x}, \bar{y} \varepsilon \bar{R}$; that is $\bar{R}$ is commutative. Since $0 \neq \bar{c} = \bar{u}\bar{v} - \bar{v}\bar{u} = 0$ we reach a contradiction. This leads us to the only

possible conclusion, namely that every commutator of R is nilpotent. Theorem 11 is thereby established.

By Lemma 4 we then have

COROLLARY. *For all* $x, y \in R$, $xy - yx \in Z$.

Since $xy - yx \in Z$, an easy induction establishes

LEMMA 12. *For all integers* $k > 1$ *and all* $x, y \in R$,

$$x^k y - y x^k = k x^{k-1}(xy - yx).$$

Suppose that $x^n - x \in Z$, $n > 1$. Then $x^n y - y x^n = xy - yx$. But we also have, from the above, that $x^n y - y x^n = n x^{n-1}(xy - yx)$. Combining these we have

THEOREM 13. *If* $x^n - x \in Z$, $n > 1$, *then for all* $y \in R$

$$n x^{n-1}(xy - yx) = (xy - yx).$$

Let $A(S) = \{x \in R \mid Sx = (0)\}$. $A(S)$ is an ideal of R, and since $S^2 = (0)$, $A(S) \supset S$; so $A(S) \neq (0)$.

THEOREM 14. $A(S) \subset Z$.

Proof. Suppose that $x \in A(S)$, $x^n - x \in Z$, $n > 1$; suppose also that for some $y \in R$, $xy - yx \neq 0$. Since

$$n x^{n-1}(xy - yx) = (xy - yx) \neq 0,$$

$R(xy - yx) \neq (0)$. Since $xy - yx \in Z$, $R(xy - yx)$ is an ideal of R, and not being (0), $R(xy - yx) \supset S$. If $s \neq 0 \in S$, then s can be written as $s = r(xy - yx)$ for some $r \in R$. Thus

$$\begin{aligned} s = r(xy - yx) &= r(n x^{n-1})(xy - yx) \\ &= nr(xy - yx)x^{n-1} = n s x^{n-1} = 0, \end{aligned}$$

since $x \in A(S)$. This proves that $xy - yx = 0$ for all $y \in R$, so $x \in Z$ and $A(S) \subset Z$.

As a consequence we also have

THEOREM 15. *If* $A(S) = R$ *then* R *is commutative.*

Since the case $A(S) = R$ is thereby settled, we *henceforth assume that* $A(S) \neq R$.

We now obtain

THEOREM 16. *All divisors of zero in* R *are in* $A(S)$.

Proof. Suppose that $xa = 0$, $x \neq 0$, $a \neq 0$. By Lemma 6 we are free to assume that x could have been taken as an element of Z. If $Rx = (0)$, then $A = \{x \in R \mid Rx = (0)\}$ would be a non-zero ideal of R and would then contain S; that is $SR = RS = (0)$, from which it follows that $A(S) = R$, a possibility we have already ruled out. So $Rx \neq (0)$. Since $x \in Z$, Rx is an ideal of R; so $Rx \supset S$. This yields $(0) = Rxa \supset Sa$. Thus $a \in A(S)$ completing the proof of Theorem 16.

THEOREM 17. *$R/A(S)$ is a field.*

For suppose that $x \in R$, $x \notin A(S)$. Since x is not a divisor of zero, $xs \neq 0$ for any $s \neq 0$ in S. Pick $s \neq 0$ in S. $Rs = sR \neq (0)$, for otherwise, as above we would obtain $SR = (0)$. Since sR is an ideal of R, and $sR \subset S$, the fact that $sR \neq (0)$ forces $sR = S$. Since $xs \neq 0 \in S$, we also have $xsR = S$. Hence for some $y \in R$, $s = xsy = xys$, since $s \in Z$. Let $e = xy$. For all $w \in R$, $(we - w)s = 0$, so by Theorem 16 $we - w \in A(S)$. Given any $w \in R$, $w \notin A(S)$ we can, using the same argument used in obtaining e, find a v so that $wv - e \in A(S)$. $R/A(S)$ is then a field with $e + A(S)$ as its unit element.

THEOREM 18. *$R/A(S)$ is of characteristic $p \neq 0$.*

Proof. Pick $x \in R$, $x \notin A(S)$. Then for some $n > 1$,

(i) $x^n - x \in Z$.

By Theorem 13, $(nx^n - x)(xy - yx) = 0$, so $nx^n - x$ is a zero divisor, forcing it to be in $A(S)$. We then have

(ii) $nx^n - x \in A(S) \subset Z$.

From (i) and (ii) we get $(n-1)x \in Z$. So for all $y \in R$, $(n-1)(xy - yx) = 0$; this also gives $((n-1)x)(xy - yx) = 0$, that is, $(n-1)x$ is a divisor of zero, and therefore must be in $A(S)$. In $R/A(S)$ this reflects into $(n-1)\bar{x} = 0$ with $\bar{x} \neq 0$ and $n - 1 \not\equiv 0$. This means that $R/A(S)$ is of characteristic $p \neq 0$.

THEOREM 19. *R is commutative.*

Proof. Let P be the prime field of $R/A(S)$. By Theorem 18 P is a finite field. Let $x, y \notin A(S)$. By Theorems 13 and 16, for suitable integers $n > 1$, $m > 1$, $n\bar{x}^{n-1} = 1$, $m\bar{y}^{m-1} = 1$. Hence $Q = P(\bar{x}, \bar{y})$, the field obtained by adjoining both $\bar{x}$ and $\bar{y}$ to P is a FINITE field. Thus in Q, $t^{p^k} = t$ for all $t \in Q$, where $k > 0$ is a FIXED integer. Let $Q^* = \{b \in R \mid b + A(S) \in Q\}$.

Q^* is clearly a subring of R, and in Q^*, $r^{p^k} - r \in A(S) \subset Z$ for all $r \in Q^*$. The main theorem of [1] consequently forces Q^* to be commutative. Since $x, y \in Q^*$, $xy = yx$. So $xy = yx$ for all $x, y \notin A(S)$. Since $xz = zx$ for $z \in A(S) \subset Z$, all $x \notin A(S)$ are in Z. Combined with $A(S) \subset Z$ we have that R must be commutative.

Between Theorems 19 and 15 all the possibilities, when R is subdirectly irreducible, are taken care of. Thus we have completely proved

THEOREM 20. *If R is a subdirectly irreducible ring with center Z, and $x^{n(x)} - x \in Z$ for every $x \in R$ with $n(x) > 1$, then R is commutative.*

Using the fact that every ring is isomorphic to a subdirect sum of subdirectly irreducible rings we are able to achieve the result which was the goal of this paper, namely

THEOREM 21. *If R is a ring with center Z such that $x^{n(x)} - x \in Z$ for all $x \in R$, where $n(x) > 1$ is an integer depending on x, then R is commutative.*

COWLES COMMISSION FOR RESEARCH IN ECONOMICS AND
THE UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

[1] I. N. Herstein, "A generalization of a theorem of Jacobson," *American Journal of Mathematics*, vol. 73 (1951), pp. 756-762.

[2] ———, "A generalization of a theorem of Jacobson II," Abstract, *Bulletin of the American Mathematical Society*, vol. 58 (1952), p. 474.

[3] N. Jacobson, "Structure theory of algebraic algebras of bounded degree," *Annals of Mathematics*, vol. 46 (1945), pp. 695-707.

[4] ———, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

[5] M. Krasner, "The non-existence of certain extensions," *ibid.*, vol. 75 (1953), pp. 112-116.

THE NON-EXISTENCE OF CERTAIN EXTENSIONS.*

By Marc Krasner.

I. N. Herstein recently asked me the following question which arose in connection with some problems in ring theory: Let K be a finite algebraic extension of a finite field k; is it possible, in any situation other than an algebraic extension of a finite field, that every $\alpha \varepsilon K$ satisfies a non-linear equation $x^{n(\alpha)} - x - a_\alpha = 0$, where $n(\alpha) > 1$ is an integer depending on α and a_α is an element of k which also depends on α. It will be proved below that the answer to this question is in the negative. This result is used by Dr. Herstein in his paper which follows the present one.[1]

The result is combined in the following

Theorem. *Let K be a finite extension of a field k. Suppose that every $\alpha \varepsilon K$ satisfies an equation of the form $x^{n(\alpha)} - x - a_\alpha = 0$ where $n(\alpha)$ is an integer larger than* 1 *and depends on α and $a_\alpha \varepsilon k$ also depends on α. Then K is an algebraic extension of a finite field.*

1. Let K/k be a finite algebraic extension of degree greater than 1 such that each $\alpha \varepsilon K$ satisfies an equation $x^{n(\alpha)} - x - a_\alpha = 0$ *with* $a_\alpha \varepsilon k$, $n(\alpha) > 1$ both depending on α. *Then if $\alpha \not\varepsilon k$ and if $\lambda \varepsilon k$, there exist only a finite number of λ-values for which $n(\lambda\alpha)$ has the same prescribed value.* Indeed, $\beta = \lambda\alpha$ satisfies an equation $y^{n(\beta)} - y - a_\beta = 0$. Hence $\alpha = \lambda^{-1}\beta$ satisfies $(\lambda x)^{n(\beta)} - \lambda x - a_\beta = 0$, i. e., $x^{n(\beta)} - \lambda^{-n(\beta)+1}x - \lambda^{-n(\beta)}a_\beta = 0$. If $n(\alpha) = n(\beta)$, we have $(\lambda^{-n(\beta)+1} - 1)\alpha = a_\alpha - \lambda^{-n(\beta)}a_\beta \varepsilon k$ and if $\lambda^{n(\beta)-1} \neq 1$, then $\alpha \varepsilon k$, contrary to hypothesis. On replacing α by any element in the set $k\alpha$, the proof is complete. *Thus, if $\alpha \not\varepsilon k$ and if $\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots$ is an infinite sequence of elements of k, then $n(\beta_i)$, where $\beta_i = \lambda_i\alpha$, converges to $+\infty$.*

2. *Reduction of the problem.* Suppose that k is not an algebraic extension of a finite field. *Then if an extension K/k of the considered kind*

* Received April 22, 1952.

[1] I. N. Herstein, "A generalization of a theorem of Jacobson III," *American Journal of Mathematics*, vol. 75 (1953), pp. 105-111.

exists, there exists such an extension, in which K has one of the following three forms: a) *pure transcendental extension $R(t)$ of the rational number field R*; b) *pure transcendental extension $P(t)$ of the field P of p elements*; c) *an algebraic number field of finite degree.*

Indeed, consider an $\alpha \varepsilon K$; then $a_\alpha = \alpha^{n(\alpha)} - \alpha$ is contained in the field $\tilde{K}$ obtained by the adjunction of α to the prime field of K (or k), and, also, in $\tilde{k} = \tilde{K} \cap k$. If $\alpha \notin k$, we have $\tilde{K} \neq \tilde{k}$. If $\beta \varepsilon \tilde{K}$, the field obtained by the adjuction of β to the prime field is contained in $\tilde{K}$; so that a $a_\beta \varepsilon \tilde{K} \cap k = \tilde{k}$. Thus $\tilde{K}/\tilde{k}$ is an extension of the considered kind.

If the characteristic of K is 0, then $\tilde{K}$ is one of the form a) or c). If the characteristic is $p \neq 0$, then, since we have supposed k not to be an algebraic extension of P, there exist in K elements which are transcendental over P. All these elements can not belong to k, for if they did, then, as each element $\beta \varepsilon K$ algebraic over P can be represented as the difference $(t + \beta) - t$ of two elements transcendental over P (and in K), we would have $K = k$. Thus we can choose as α an element transcendental over P (not in k) and $\tilde{K}$ is of the form b).

3. Let $|\cdot\cdot\cdot|$ be a valuation of K, and let $\bar{K}$, $\bar{k}$ be the completions of K, k with respect to this valuation. Suppose that *there exists a compact subgroup $\bar{U}^*$ of the multiplicative group $\bar{U}$ of units of $\bar{k}$ (for the considered valuation) such that $U^* = \bar{U}^* \cap k$ is an infinite set. Then if K/k is an extension of the considered kind, we have $\bar{K} = \bar{k}$.*

Let ξ be an element of K. On multiplying ξ by a convenient $\mu \neq 0 \varepsilon k$ we can obtain an $\alpha = \mu\xi$ such that $|\alpha| < 1$. Let $\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots$ be an infinite sequence of elements of U^* and let $\beta_i = \lambda_i\alpha$. Then for each i, $\alpha^{n(\beta_i)} - \lambda_i^{-n(\beta_i)+1}\alpha - c_{\alpha,i} = 0$, where $c_{\alpha,i} = \lambda_i^{-n(\beta_i)}a_{\beta_i} \varepsilon k$. But $n(\beta_i) \rightarrow +\infty$ and since $|\alpha| < 1$, $|\alpha^{n(\beta_i)}| \rightarrow 0$. On the other hand, as $\lambda_i^{-n(\beta_i)+1} \varepsilon U^*$ forms an infinite sequence in the compact set $\bar{U}^*$ of $\bar{k}$, this sequence has at least one limit point $\bar{\lambda} \varepsilon \bar{U}^* \subseteq \bar{k}$. Suppose that everything has been renumbered so that a subsequence converging to $\bar{\lambda}$ is given by $\{\lambda_i^{-n(\beta_i)+1}\}$, i. e., $\lambda_i^{-n(\beta_i)+1} \rightarrow \bar{\lambda}$. Then also the sequence $a_{\alpha,i} = \alpha^{n(\beta_i)} - \lambda_i^{-n(\beta_i)+1}\alpha \rightarrow -\bar{\lambda}\alpha$ converges and its limit is in $\bar{k}$. Hence $\bar{\lambda}\alpha \varepsilon \bar{k}$, and as $\bar{\lambda} \varepsilon \bar{k}$, we have $\alpha \varepsilon \bar{k}$ and $\xi = \mu^{-1}\alpha \varepsilon \bar{k}$. This proves that $\bar{K} = \bar{k}$.

4. The cases a) and b), K a pure transcendental extension $C(t)$ of transcendence degree 1 over a prime field C ($= R$ or P). Then, by Lüroth's

theorem, k, which certainly contains the prime field C, is of the form $C(\theta)$, *where* $\theta = \phi(t)$ *is a polynomial in* t *with coefficients in* C. As for each $a \varepsilon C$, $a \neq 0$, $C(ax) = C(x)$ and for each $A \varepsilon C$, $A \neq 0$, $C(A\theta) = C(\theta)$, we can, without changing K/k, replace $\phi(t)$ by any $A\phi(ax)$, $a \neq 0$, $A \neq 0$ and $a, A \varepsilon C$.

5. In the case of a) we shall take as $|\cdots|$ a valuation extending some p-adic valuation $|\cdots|_p$ of R. Before this extension however we multiply t and θ by elements of R in such a way that $\phi(t)$ will be replaced by a normed polynomial $\phi(t) = t^m + u_1 t^{m-1} + \cdots + u_m$ in which all its terms except the first are divisible by p (which is clearly possible). When that is done, define the valuation of K by putting $|\xi| = \operatorname{Max}_i |a_i|_p$ for each polynomial $\xi = \sum_i a_i t^i$ $(a_i \varepsilon R)$ in t with coefficients in R.

$\bar{R}$, which is the rational p-adic field, is locally compact with respect to $|\cdots| = |\cdots|_p$, and so the group of its units $\bar{U}^*$ (which is a subgroup of the group of units $\bar{U}$ of $\bar{k} = \bar{R}(\theta)$), being bounded, is compact.

$$U^* = \bar{U}^* \cap k = \bar{U}^* \cap R,$$

which is the group of p-units of R, is infinite. Hence we must have $\bar{K} = \bar{k}$.

It will first be shown that the valuation induced by $|\cdots|$ on $k = R(\theta)$ is of the same kind, i. e., that if $\eta = \Sigma a_i \theta^i$ is a polynomial in θ, we have $|\eta| = \operatorname{Max}_i |a_i|_p = \operatorname{Max}_i |a_i|$. Indeed, by the definition of $\theta = \phi(t)$,

$$|\theta| = \operatorname{Max}_i(|1|, |u_i|) = |1|$$

(because $|u_i| < 1$ for each i) and

$$|\theta - t^m| = |\sum_{i=1}^{i=m} u_i t^{m-i}| = \operatorname{Max}_i |u_i| < 1.$$

Hence

$$|\eta - \Sigma a_i t^{im}| = |\Sigma a_i(\theta^i - t^{mi})| \leqq \operatorname{Max}_i (|a_i| \cdot |\theta - t^m|)$$
$$< \operatorname{Max}_i |a_i| \text{ and } |\Sigma a_i t^{mi}| = \operatorname{Max}_i |a_i|.$$

Consequently

$$|\eta| = |\Sigma a_i t^{im}| = \operatorname{Max}_i |a_i|.$$

If $\bar{K} = \bar{k}$, then $t \varepsilon \bar{k}$. Hence there exists an element $\xi \varepsilon k$ such that in $\bar{K}$, and therefore also in K (as the valuation extends that of K and t and

ξ are both in K), we must mave $|t-\xi|<1$. Since $|t|=1$, we have in particular $|\xi|=1$. But ξ is a quotient $g(\theta)/h(\theta)$ of two polynomials in θ, and we can suppose that the coefficients of $h(\theta)$ are all integers in R and not all divisible by p. Then $|h(\theta)|=1$, and we have $|th(\theta)-g(\theta)|<1$, that is, all the coefficients of $th(\theta)-g(\theta)$ are p-integers of R divisible by p. But $h(\theta)\equiv h(t^m) \pmod p$; hence $th(t^m)\equiv g(t^m) \pmod p$, where $h(\theta)$ has at least one coefficient not divisible by p. Then $th(t^m)$ has at least one term with coefficient $\not\equiv 0 \pmod p$ and with degree $\equiv 1 \pmod m$, and the degrees of all the terms of $g(t^m)$ are $\equiv 0 \pmod m$. Hence $m=1$ and $K=k$, against the hypothesis.

6. In the cases b) and c) we consider any non-archimedean valuation of K. Then, as $\bar{K}$ is locally compact, we can take as $\bar{U}^*$ the whole group of units $\bar{U}$ of $\bar{k}$, and $\bar{U}\cap k$ is clearly infinite. Hence $\bar{K}=\bar{k}$ holds again.

It follows that *any prime ideal* $\mathfrak{p}$ *of* k *is completely decomposed in* K; i. e., each of its prime factors $\mathfrak{P}$ in K has its residual degree f in K/k and its ramification order e in K/k both equal to 1.

7. In case b) consider the ideal $\mathfrak{p}_a=(\theta-a)$, $a\,\varepsilon\, P$, of k. The residual field, r_a, of $\mathfrak{p}_a$ in k is P, the prime field. Hence if $\mathfrak{P}$ is a prime factor of $\mathfrak{p}_a$ in K, there must be, as $|t|=1$, some $b\,\varepsilon\, P$ such that $t\equiv b \pmod{\mathfrak{P}}$. Thus $\mathfrak{P}=(t-b)$, and as $\mathfrak{p}_a=(\theta-a)=(\phi(t)-a)$, we must have $\phi(t)-a\equiv 0 \pmod{(t-b)}$. But it cannot be that $\phi(t)-a\equiv 0 \pmod{(t-b)^2}$, for then $\mathfrak{p}_a$ would be divisible by $\mathfrak{P}^2$, hence $e\neq 1$. Consequently, if m is the degree of $\phi(t)$, *$\phi(t)-a$ must decompose in m linear factors in P, for each $a\,\varepsilon\, P$.* But if $a\neq a'$, $a, a'\,\varepsilon\, P$, $(\phi(t)-a)-(\phi(t)-a')=a-a'\neq 0$, and $\phi(t)-a$ and $\phi(t)-a'$ can have no common factor. Hence $\prod_{a\varepsilon P}(\phi(t)-a)$ must decompose in mp different linear factors in the field P of p elements. This is only possible if $mp=p$, $m=1$; i. e., if $K=k$, which violates hypothesis.

8. There remains only the case c). Consider the Galois field (normal extension) K^* of K/k; that is, the field composed by the conjugate fields of K over k. If each prime ideal of k is completely decomposable in K, it is also completely decomposable in K^*. Let $\mathfrak{G}$ be the Galois group of K^*/k and let $\mathfrak{g}$ be some Abelian subgroup of $\mathfrak{G}$ other than the identity. Let k^* be a subfield of K^* belonging to $\mathfrak{g}$. Then each prime ideal of k^* is completely decomposable in K^*, and K^*/k^* is an Abelian extension such that $K^*\neq k^*$.

But a prime ideal $\mathfrak{p}^*$ of k^* is completely decomposable in K^* if and only if it belongs to the Takagi group H_{K^*/k^*}, and if $K^* \neq k^*$, there exist prime ideals $\mathfrak{p}^*$ of k^* which do not belong to H_{K^*/k^*}. Thus, contrary to what has been proved above, not all ideals $\mathfrak{p}^*$ of k^* are completely decomposable. This completes case c) and the proof of the theorem.

9. The reasoning leading to $K = k$ is not applicable to the case in which K is an algebraic extension of P, because we cannot find a valuation of K satisfying the requirements of Section **3**. Indeed, in this case the only valuation of K is its trivial valuation $|\,0\,| = 0$, $|\,\alpha\,| = 1$ if $\alpha \neq 0$. But then $\bar{k} = k$ and the only compact set of $\bar{k}$ are finite sets of k.

NOTRE DAME UNIVERSITY AND
NATIONAL CENTER OF SCIENTIFIC RESEARCH, PARIS.

ON THE BEHAVIOR OF THE SOLUTIONS OF REAL BINARY DIFFERENTIAL SYSTEMS AT SINGULAR POINTS.*

By Philip Hartman and Aurel Wintner.

1. In a vicinity of $(x, y) = (0, 0)$, let

(1) $\quad f = f(x, y), \quad g = g(x, y),$ where $f(0,0) = 0, \quad g(0,0) = 0,$

be continuous functions. Consider the system of differential equations

(2) $$x' = f(x, y), \qquad y' = g(x, y), \qquad (' = d/dt),$$

and suppose that (2) has a solution

(3) $$\Gamma\colon \ x = x(t),\ y = y(t), \text{ where } 0 \leqq t < b \ (\leqq \infty),$$

satisfying

(4) $$(0,0) \neq (x, y) \rightarrow (0, 0) \text{ as } t \rightarrow b.$$

Then Γ will be referred to as a *solution path of* (2) *reaching the origin.* (There is no loss of generality in assuming that $(x(t), y(t))$ tends to the origin for *increasing* t, since otherwise f, g would be replaced by $-f$, $-g$). Corresponding to (4), let $r = r(t) > 0$ and $\theta = \theta(t)$ be continuous functions defined by

(5) $$r^2 = x^2 + y^2 \text{ and } \theta = \operatorname{arc\,tan} y/x$$

(and say $0 \leqq \theta(0) < 2\pi$).

This paper will be concerned with the question of the existence of a tangent to the solution path Γ at $(0, 0)$, that is, with the question whether or not

(6) $$\theta_0 = \lim_{t \to b} \theta(t) \text{ exists}, \qquad (-\infty < \theta_0 < \infty),$$

and, incidentally, with the question whether or not (6) implies that

(7) $$\lim_{t \to b} \psi(t) = \theta_0 \ (\operatorname{mod} \pi), \text{ where } \psi(t) = \operatorname{arc\,tan} y'(t)/x'(t).$$

It is not supposed that $f^2 + g^2 > 0$ if $x^2 + y^2 > 0$, that is, that $(x, y) = (0, 0)$ is an isolated singular point of (2). It will be understood that (7) will be meant so as to imply that $x'(t)$ and $y'(t)$ do not vanish simultaneously for any t near b (so that $\psi(t)$ is defined mod π).

* Received May 7, 1952.

In § 2-§ 6, general theorems will be developed. In § 7-§ 8, these will be applied to cases in which (2) is of the form $x' = \alpha x + \beta y + o(r)$, $y' = \gamma x + \delta y + o(r)$, as $r \to 0$, where $\alpha\delta - \beta\gamma \neq 0$.

2. It will be convenient to formulate a set of conditions (†) to be assumed below for the pair of functions f, g in (2).

(†) Let $f(x, y)$, $g(x, y)$ be continuous functions on some circle $x^2 + y^2 \leqq$ const. Let Ω^* be a closed periodic set, of period 2π, which contains no interval. Let the limits

$$F(\theta) = \lim_{r \to +0} f(r \cos\theta, r \sin\theta)/r, \qquad G(\theta) = \lim_{r \to +0} g(r \cos\theta, r \sin\theta)/r \tag{8}$$

exist uniformly on every closed θ-interval not containing any point of Ω^*, and let the functions (8) have the properties that

$$F^2(\theta) + G^2(\theta) > 0 \text{ for } \theta \text{ not in } \Omega^* \tag{9}$$

and that the set Ω^0 of θ-values not contained in Ω^* and satisfying the equation

$$J(\theta) = 0, \tag{10}$$

where

$$J(\theta) = G(\theta) \cos\theta - F(\theta) \sin\theta, \tag{11}$$

is either empty or an infinite sequence which has no finite cluster value in the complement of Ω^*.

It is to be noted that these conditions (†) on the functions f, g do not imply that a solution of (2) is determined uniquely by initial conditions, since nothing like a Lipschitz condition is assumed.

Let the function $K(\theta)$ be defined for θ not in Ω^* by

$$K(\theta) = G(\theta) \sin(\theta) + F(\theta) \cos\theta. \tag{12}$$

Then, by (9),

$$J^2(\theta) + K^2(\theta) \neq 0 \text{ for } \theta \text{ not in } \Omega^*. \tag{13}$$

The functions F, G, J, K are periodic functions of period 2π (although they are considered to be undefined on Ω^*). Correspondingly, Ω^0 and Ω^* are periodic sets of period 2π.

Condition (†) is invariant under rotations of the (x, y)-plane about the origin and under changes of scale on the axes of x, y, t.

Under the assumption (†), the following statements will be proved:

(i) *If* Γ *is a solution path* (3) *of* (2) *reaching the origin and satisfying* (6), *then* θ_0 *is either in* Ω^* *or in* Ω^0; *in the latter case* (7) *holds.*

On the other hand, (6) does not imply (7) if θ is in Ω^* (even when $(x,y)=(0,0)$ is an isolated singular point of (2)). This will be proved by an example.

(ii) *If* Γ *is a solution path* (3) *of* (2) *reaching the origin and if* (6) *does not hold, then* Γ *is a spiral, that is,*

$$|\theta(t)|\to\infty \text{ as } t\to b. \tag{14}$$

Hence $\lim\theta(t)$, as $t\to b$, always exists if ∞ or $-\infty$ is allowed as a limit.

(iii) *If* $J(\theta)$ *changes sign at a point* $\theta=\theta_0$ *of* Ω^0, *then* (2) *has at least one solution path* (3) *reaching the origin and satisfying* (6).

(iv) *If* $J(\theta)$ *assumes both positive and negative values, then every solution path* (3) *of* (2) *reaching the origin satisfies* (6).

The proofs of (i)-(iv) will depend on the methods of [7] and the results of [3].

3. *Proof of* (i). The functions (5) of t, where $x=x(t)$, $y=y(t)$, satisfy the differential equations

$$\begin{aligned} r' &= g(r\cos\theta, r\sin\theta)\sin\theta+f(r\cos\theta,r\sin\theta)\cos\theta,\\ r\theta' &= g(r\cos\theta,r\sin\theta)\cos\theta-f(r\cos\theta,r\sin\theta)\sin\theta. \end{aligned} \tag{15}$$

If θ is not near a point of Ω^*, then (15) can be written as

$$r^{-1}r'=K(\theta)+o(1),\qquad \theta'=J(\theta)+o(1)\qquad\qquad (r\to 0). \tag{16}$$

Let (4) and (6) hold for a θ_0 not in Ω^*. It will be shown that θ_0 is in Ω^0. Suppose the contrary. Then $J(\theta)=J(\theta(t))$ exists and is not 0 for t near b. Hence $\theta'\neq 0$, by (16), and so θ can be introduced as an independent variable. Then $r=r(\theta)$ is defined on some interval $\theta_0-\delta\leqq\theta<\theta_0$ or $\theta_0<\theta\leqq\theta_0-\delta$, where $\delta>0$, and satisfies

$$dr/d\theta=r(K(\theta)+o(1))/(J(\theta)+o(1)) \tag{17}$$

and $r(\theta)\to 0$, as $\theta\to\theta_0$. But this is contradictory, since $dr/d\theta=O(r)$ implies that $r(\theta)>0$ cannot reach 0 for a finite value of $\theta=\theta_0$. This proves that θ_0 is in Ω^0 when it is not in Ω^*.

It remains to show that when θ is in Ω^0, then (6) implies (7). There is no loss of generality in subjecting the (x,y)-plane to a rotation and then supposing that $\theta_0=0$. Then, by (10) and (11), $G(0)=0$ and, by (9),

$F(0) \neq 0$. Hence, by the uniformity of the limits (8), $x'/r = f/r \neq 0$ for t near b, and

$$(18) \qquad y'/x' = [g(r\cos\theta, r\sin\theta)/r]/[f(r\sin\theta, r\cos\theta)/r] \to G(0)/F(0) = 0,$$

as $t \to b$. This proves (7) and completes the proof of (i).

As to the remark following (i), let $f(x,y) = -x$ and let $g(x,y) = -2x^2\sin 1/x + x\cos 1/x$ for $0 < |y| \leqq x^2$. Since (†) involves, in the main, conditions on $g(x,y)$ along lines $\theta =$ const., it is clear that the definition of this g can be extended to a circle $x^2 + y^2 \leqq$ Const. so as to satisfy (†), where $\theta = 0 (\mathrm{mod}\, 2\pi)$ is the only point of Ω^*. But for any extension of g, (2) has the solution $x = e^{-t}$, $y = e^{-2t}\sin e^t$. This solution reaches the origin (as $t \to \infty$) and satisfies (6), with $\theta_0 = 0$, but (7) fails to hold.

4. *Proof of* (iii). If θ is near θ_0, then $K(\theta) \neq 0$, hence the second equation in (16) can be divided by the first when r is sufficiently small. There results the differential equation

$$(19) \qquad r d\theta/dr = (J(\theta) + o(1))/(K(\theta) + o(1)),$$

where the right-hand side is a continuous functions of (r, θ), say $R(r,\theta)$ (when θ is near θ_0 and $r \geqq 0$ is small). Furthermore, if $\delta > 0$ is sufficiently small, then $R(r, \theta_0 + \delta)$ and $R(r, \theta_0 - \delta)$ are of opposite sign for small r, and $R(0,\theta)$ is not 0 on the interval $|\theta - \theta_0| \leqq \delta$ except at $\theta = \theta_0$. Theorems (I), (II) in [3], p. 304, show that (19) has a solution $\theta = \theta(r)$, defined on an interval $0 < r \leqq \epsilon$ and satisfying $\theta(r) \to \theta_0$ as $r \to 0$. If $\theta = \theta(r)$, the first equation of (16) determines t as a function of r. This function of t is monotone with a non-vanishing derivative, and possesses therefore an inverse $r = r(t)$ having a continuous derivative. The function $\theta = \theta(r(t))$ satisfies the second equation of (16). Clearly, $r(t)$, $\theta(r(t))$ supply, by virtue of (5), a solution path (3) of (2) reaching the origin and satisfying (6).

5. *Proof of* (iv). Suppose that (3) is a solution path of (2) which reaches the origin but fails to satisfy (6). Then, as $t \to b$, the function $\theta = \theta(t)$ either has no finite cluster value or has at least one cluster value not in Ω^* (since Ω^* contains no intervals). Hence, in either case, Ω^* can be covered with a periodic, non-empty, open set Θ, of period 2π, with the properties that Θ consists of a finite number of intervals on a period, that $J(\theta) \neq 0$ for any θ which is an endpoint of such an interval, and that $\theta = \theta(t)$ is not in Θ for certain t arbitrarily near b.

The set of points θ not in $\mathfrak{S}$ can be divided into a finite number of periodic, pair-wise disjoint sets $\Sigma_1, \cdots, \Sigma_N$, where Σ_i is obtained by choosing a closed interval in the complement of $\mathfrak{S}$ and letting Σ_i contain this interval and all of its translations by $2n\pi$, where $n = 0, \pm 1, \cdots$.

At least one of the sets $\Sigma_1, \cdots, \Sigma_N$, say Σ, contains a sequence of points $\theta = \theta(t_n)$, where $t_n < t_{n+1} \rightarrow b$, as $n \rightarrow \infty$.

Suppose first that $\theta(t)$ remains within one interval I: $\theta_1 \leqq \theta \leqq \theta_2$ of Σ for all t near b. Then (16) holds for all t near b. Also, $\theta(t)$ is bounded but has no limit, by assumption. Hence $\theta(t)$ is not monotone. Consequently, $\theta'(t)$ changes sign an infinity of times. Thus I contains a non-empty, finite set of points θ at each of which points the function $J(\theta)$ changes sign, and there must exist at least one of these points, say $\theta = \theta_0$, having the property that $\theta(t^n) = \theta_0$ for a sequence of t-values $t^1, t^2, \cdots$ satisfying $t^n \rightarrow b$ as $n \rightarrow \infty$. On the other hand, if $J(\theta) \neq 0$, then $\theta(t) = \theta$ cannot hold for more than one t-value near b, since (16) assures that $\theta(t)$ has the same sign as $J(\theta)$ if t is sufficiently near b. But these facts imply that (6) holds, which is a contradiction.

Thus, in order to complete the proof of (iv), there remains to consider the case in which $\theta(t_n)$ is in Σ for a sequence of t-values t_n which tend to b, but $\theta(t)$ does not stay in one and the same interval I of Σ for all t near b. Let Σ consist of the intervals

$$I_n\colon\ \theta_1 + 2n\pi \leqq \theta \leqq \theta_2 + 2n\pi, \text{ where } n = 0, \pm 1, \cdots.$$

First, the case $J(\theta_2)J(\theta_1) < 0$ is impossible. In order to see this, suppose that $J(\theta_2) < 0$ and $J(\theta_1) > 0$. Then $\theta(t)$ remains in one and the same I_n for large t, since $\theta'(t)$ would become negative [or positive] if $\theta(t)$ could assume the value θ_2 [or θ_1] $\operatorname{mod} 2\pi$. On the other hand, if $J(\theta_2) > 0$ and $J(\theta_1) < 0$, and if $\theta(t)$ is not in an I_n for some t near b, then it cannot enter an I_n for larger values of t.

Finally, the case $J(\theta_2)J(\theta_1) > 0$ is impossible. For suppose that $J(\theta_2) > 0$ and $J(\theta_1) > 0$. Then, as t increases to b, $\theta(t)$ can enter I_n only by assuming the value θ_1 $(\operatorname{mod} 2\pi)$, and then leave I_n by attaining the value θ_2 $(\operatorname{mod} 2\pi)$. Hence $\theta(t)$ passes successively through $I_n, I_{n+1}, \cdots$. In particular, $\theta(t)$ assumes, at least once, every value exceeding $\theta_1 + 2m\pi$, where m is a fixed large integer. Since $J(\theta)$ attains both positive and negative values, there must exist points θ^1, θ^2 not in Ω^* such that $\theta^1 < \theta^2$ and $J(\theta^1)J(\theta^2) < 0$. Since $\theta(t)$ passes (at least once) through each of the intervals $\theta^1 + 2n\pi \leqq \theta \leqq \theta^2 + 2n\pi$ for large n, this leads to the contradiction obtained in the last paragraph. The case in which $J(\theta_1) < 0$ and $J(\theta_2) < 0$ is treated similarly. This proves (iv).

6. *Proof of* (ii). The assertion (ii) is a consequence of (iv) and its proof. Using the notation of the preceding proof, it is seen that if $\theta(t)$ has no infinite limit, then (for a suitable choice of Θ) the value of $\theta(t)$ must be in one of the sets $\Sigma_1, \cdots, \Sigma_N$, say in Σ, for an infinity of values of t near b. By the above considerations, $\theta(t)$ cannot remain in one interval I_n of Σ, nor can $J(\theta_1)J(\theta_2) < 0$ hold. Hence $J(\theta_1)J(\theta_2) > 0$, and $\theta(t)$ enters successively $I_n, I_{n+1}, \cdots$ or $I_n, I_{n-1}, \cdots$. Thus (ii) is proved.

7. Consider (2) in the case

$$x' = \alpha x + \beta y + f^*(x, y), \qquad y' = \gamma x + \delta y + g^*(x, y), \tag{20}$$

where

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \text{ is non-singular} \tag{21}$$

and f^*, g^* are continuous functions of (x, y) near $(x, y) = (0, 0)$ and satisfy

$$f^* = o(r), \; g^* = o(r), \text{ as } r^2 = x^2 + y^2 \to 0. \tag{22}$$

In the analytic case, the discussion of (20) goes back to Poincaré, and in the non-analytic case to Bendixson [1] and Perron [4], [5].

After an affine transformation of the (x, y)-plane and a change of scale and direction on the t-axes, it can be supposed that A has one of the normal forms

$$(23_1)\ \begin{pmatrix} -1 & 0 \\ 0 & \delta \end{pmatrix}; \qquad (23_2)\ \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix};$$

$$(23_3)\ \begin{pmatrix} -1 & \beta \\ -\beta & -1 \end{pmatrix}; \qquad (23_4)\ \begin{pmatrix} 0 & \beta \\ -\beta & 0 \end{pmatrix}.$$

It is known that, in the cases (23_1), (23_2), (23_3), condition (22) is sufficient to assure that, in some sense, the solution paths of (26) behave like those of the corresponding linear system, where $f^* \equiv g^* \equiv 0$. For example, in the cases (23_2), (23_3) and (23_1) with $\delta < 0$, a solution path $x = x(t)$, $y = y(t)$ of (20), passing through a point (x_0, y_0) sufficiently near $(0, 0)$, can be continued for all larger values of t, and all such continuations reach the origin as $t \to \infty$; cf. [8], p. 816. In the case (23_1) with $\delta > 0$, there exist some solution paths (3) of (20) reaching the origin as $t \to \infty$, but every sufficiently small circle C_ϵ: $x^2 + y^2 < \epsilon^2$ contains points (x_0, y_0) such that any solution path through (x_0, y_0) leaves the circle C_ϵ (for increasing as well as for decreasing t); cf. the proof of (II) in [3], p. 306.

In the case (23_3), all solution paths Γ reaching the origin are spirals, that is, they satisfy (14), where $b = \infty$; cf. (iib) in [8], p. 817. In the

case (23_1) with $\delta = -1$, it is known that a solution path reaching the origin need not have a tangent there, since it can be a spiral (cf. [1], pp. 128-129 or [8], p. 818), although this does not happen when $f^* \equiv g^* \equiv 0$. The same choice of f^*, g^* as in the example just referred to shows that this phenomenon can occur in the case (23_2) also. On the other hand, it cannot occur in the case $\delta = -1$ of (23_1) if f^*, g^* are subject to a certain condition more severe than (22); cf. [8], p. 823.

There does not seem to be known any general theorem concerning the existence of tangents at the origin for solution paths of (20) in the case (23_1) if $\delta \neq 0, -1$. (In this direction, cf. [6], pp. 209-210, for the analytic case; Satz 3, p. 129 and Satz 2, p. 271 in [5]; and Corollaries 3, 4 in [2], pp. 501-502.) Such a theorem can, however, be deduced from (i)-(iv), as follows:

(a) *In the case* $\delta \neq 0, -1$ *of* (23_1), *every solution path of* (20)-(22) *reaching the origin has a tangent there, and the latter is the limit of tangents. Furthermore, every such solution is tangent to a solution path of the corresponding linear system* (*where* $f^* \equiv g^* \equiv 0$), *and conversely.*

In the case of a multiple elementary divisor, only a weaker statement holds:

(b) *In the case* (23_2) *of* (20)-(22), *every solution path reaching the origin either is tangent to the* y*-axis or satisfies* (14).

There are cases (23_2) of (20)-(22) realizing any of the following possibilities: All, some but not all, no solution paths reaching the origin are tangent to the y-axis there.

For the case of a multiple characteristic number but simple elementary divisors, the statement to be made is entirely negative:

(c) *In the case* $\delta = -1$ *of* (23_1), *a solution path* (20)-(22) *reaching the origin can satisfy any of seven possibilities compatible with*

$$-\infty \leqq \liminf_{t \to \infty} \theta(t) \leqq \limsup_{t \to \infty} \theta(t) \leqq \infty. \tag{24}$$

8. *Proof of* (a). In any of the cases (23), the limits (8) exist uniformly for all θ. In the case (23_1), $F(\theta) = -\cos\theta$, $G(\theta) = \delta \sin\theta$ and $J(\theta) = (\delta + 1)\cos\theta \sin\theta$, $K(\theta) = -\cos^2\theta + \delta \sin^2\theta$, hence (9) is satisfied. If $\delta = -1$, then $J(\theta) \equiv 0$, and (†) does not hold. If $\delta \neq -1$, then the set Ω^0 consists of the points $n\pi/2$, where $n = 0, \pm 1, \cdots$, and (†) is fulfilled. Since $J(\theta)$ changes sign at every point $\theta = \theta_0$ of Ω^0, the assertion (a) follows from (ii) and (iv).

Proof of (b). In the case (23_2), $F(\theta) = -\cos\theta$, $G(\theta) = \cos\theta - \sin\theta$ and $J(\theta) = \cos^2\theta$, $K(\theta) = 1 - \sin\theta\cos\theta$. Hence (9) is satisfied, while $J(\theta) = 0$ holds if and only if $\theta = (2n+1)\pi/2$ for $n = 0, \pm 1, \cdots$. Thus (†) is fulfilled and (b) follows from (i) and (ii).

The example in [5], pp. 128-129 or [8], p. 818, can be modified to prove the remark following (b). Let

$$f^* = -rh(r)k(\theta)\sin\theta \text{ and } g^* = rh(r)k(\theta)\cos\theta,$$

where $h(r)$, $k(\theta)$ will be specified below. The function $k(\theta)$ will be continous and of period 2π. The function $h(r)$ will be defined and continuous for small $r > 0$ and will satisfy $h(r) \to 0$, so that (22) holds. In polar coordinates, the system (20) becomes

$$(25) \qquad r' = -r(1 - \tfrac{1}{2}\sin 2\theta), \qquad \theta' = \cos^2\theta + h(r)k(\theta).$$

When $h(r) \equiv 0$, so that $f^* \equiv g^* \equiv 0$, all solution paths of (20) reach the origin (as $t \to \infty$) in the case (23_2) and are tangent to the y-axis at the origin.

Let $h(r) = |\log r|^{-1}$ and $k(\theta) \equiv 1$. Then $r' \geqq -\frac{1}{2}r$ and $\theta' \geqq h(r)$. Hence there exist two positive constants satisfying $r \leqq \text{const.}\, e^{-\frac{1}{2}t}$ and $\theta'(t) \geqq \text{Const.}\, t^{-1}$ for any solution passing through a point $(r_0, \theta_0) \neq (0, 0)$. Thus, in this case, every solution path reaches the origin (as $t \to \infty$) and is a spiral.

Let $h(r) = |\log r|^{-1}$ and $k(\theta) = |\cos\theta|^{\frac{1}{2}} \geqq 0$. Then (25) has the solution $r = e^{-t}$ and $\theta \equiv \frac{1}{2}\pi$, which reaches the origin and is tangent to the y-axis there. It remains to show that (25) also has spiral solutions reaching the origin. To this end, notice that a point (r_0, θ_0), where $r_0 > 0$ and $(2n-1)\pi/2 < \theta < (2n+1)\pi/2$, determines (locally) a unique solution $r = r(t)$, $\theta = \theta(t)$, which reduces to r_0, θ_0 for a given value of $t = t_0$. For this solution, $\theta(t)$ is non-decreasing and attains the value $(2n+1)\pi/2$ for a finite $t = t_1 (> t_0)$ (with $r(t_1) > 0$). In fact, $-2r \leqq r' \leqq -\frac{1}{2}r$, hence $r_0 e^{-2(t-t_0)} \leqq r(t) \leqq r_0 e^{-\frac{1}{2}(t-t_0)}$ and $\theta' \geqq h(r)k(\theta) \geqq |\cos\theta|^{\frac{1}{2}}/|\log(r_0 e^{-\frac{1}{2}(t-t_0)})|$. Since the differential equation $\theta' = |\cos\theta|^{\frac{1}{2}}/|\log(r_0 e^{-\frac{1}{2}(t-t_0)})|$ has a solution which, for $t = t_0$, has the value θ_0 and, for some finite t^* $(> t_0)$, assumes the value $(2n+1)\pi/2$, it follows that $\theta(t_1) = (2n+1)\pi/2$ for some t_1, where $t_0 < t_1 < t^*$. These considerations show that some, but not all, solutions of (25) reaching the origin are tangent to the y-axis there.

Proof of (c). Let $f^* = -rh(r)\sin\theta$ and $g^* = rh(r)\cos\theta$, where $h(r)$ is defined and continuous for $r > 0$ and satisfies $h(r) \to 0$ as $r \to 0$.

In the case $\delta = -1$ of (23_1), introduction of polar coordinates into (20) gives $r' = -r$ and $\theta' = h(r)$. This system has the solution $r = e^{-t}$, $\theta = \int^t h(e^{-u})\,du$, which, since $r \to 0$ as $t \to \infty$, is a solution path Γ reaching the origin. On the other hand, it is clear that h can be chosen so as to realize any of the seven possibilities in (24). This proves (c).

Appendix.

A classical general statement on (20)-(22), where the continuous functions f^*, g^* are arbitrary, is that there exist in general solution paths $x = x(t)$, $y = y(t)$ which tend to the origin either as $t \to \infty$ or as $t \to -\infty$, the only possible exception being the case in which A has a pair of purely imaginary characteristic numbers. Thus it is natural to inquire into conditions under which a system (2), or more generally a system

$$x' = f(t, x), \tag{25}$$

where x and f are vectors with n real components, will not possess any solution $x = x(t)$ satisfying $x(t) \to 0$ when either $t \to \infty$ or $t \to -\infty$, without being identically 0 from a certain $t = t_0$ onward (that is, for $t_0 \leqq t < \infty$ or $-\infty < t \leqq t_0$ in the *respective* cases).

It is easy to see that a sufficient condition for this situation is that $f(t, x)$ be continuous on the (t, x)-space and that there exists for small $s > 0$ a positive, continuous function $\phi(s)$ satisfying

$$\int_{+0} ds/\phi(s) < \infty \tag{26}$$

and

$$|\, x \cdot f(t, x) \,| \geqq \phi(s) > 0, \text{ where } s = x \cdot x \tag{27}$$

(the dot denotes scalar multiplication).

In order to prove this criterion, suppose that it is false. Then it can be assumed that there exists on a half-line $t_0 \leqq t < \infty$ a solution satisfying $0 \neq x(t) \to 0$ as $t \to \infty$. Put, along this solution, $s = x \cdot x$. Then (25) shows that $s = s(t)$ has the derivative $s' = 2x \cdot f$. It follows therefore from (27) that $s(t)$ is a monotone function. Consequently, if u is sufficiently large and v is greater than u, then

$$\int_u^v |\, ds(t) \,| / \phi(s(t)) \geqq \int_u^v dt. \tag{28}$$

In view of (26), this leads to a contradiction as $v \to \infty$, since, by assumption, $s(t) \to 0$ as $t \to \infty$.

It is clear from this proof that (27) can be generalized to

$$(29) \qquad | x \cdot f(t, x) | \geqq \psi(| t |)\phi(s) > 0, \qquad (s = x \cdot x),$$

where $\phi(s)$ satisfies the same conditions as above and $\psi(t)$, $-\infty < t < \infty$, is any positive continuous function of t satisfying

$$(30) \qquad \int^{\infty} \psi(t)\,dt = \infty.$$

This criterion is the dual of a result of [9], pp. 557-558, which concerns itself with the situation in which *every* (instead of *no*) solution vector $x = x(t)$ of (25) is asymptotic to the point $x = 0$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] I. Bendixson, "Sur les courbes définies par des équations différentielles," *Acta Mathematica*, vol. 24 (1901), pp. 1-88.

[2] P. Hartman, "On the solutions of an ordinary differential equation near a singular point," *American Journal of Mathematics*, vol. 68 (1946), pp. 495-504.

[3] ——— and A. Wintner, "On the asymptotic behavior of the solutions of a non-linear differential equation," *ibid.*, pp. 301-308.

[4] O. Perron, "Beweis für die Existenz von Integralen einer gewöhnlichen Differentialgleichung in der Umgebung einer Unstetigkeitsstelle," *Mathematische Annalen*, vol. 75 (1914), pp. 255-273.

[5] ———, "Ueber die Gestalt der Integralkurven einer Differentialgleichung erster Ordnung in der Umgebung eines singulären Punktes," *Mathematische Zeitschrift*, vol. 15 (1922), pp. 121-146 and vol. 16 (1923), pp. 273-295.

[6] E. Picard, *Traité d'Analyse*, 3rd edition (1928), vol. III.

[7] A. Wintner, "Unrestricted Riccatian solution fields," *Quarterly Journal of Mathematics* (Oxford), vol. 18 (1947), pp. 65-71.

[8] ———, "Vortices and nodes," *American Journal of Mathematics*, vol. 69 (1947), pp. 815-824.

[9] ———, "Asymptotic integration constants," *ibid.*, vol. 68 (1946), pp. 553-559.

ON THE CURVATURES OF A SURFACE.*

By PHILIP HARTMAN and AUREL WINTNER.

1. *The smoothness of H and K.* A point set S in the (x, y, z)-space is said to be a (small piece of a) surface of class C^n, where $n \geqq 1$, if it is the locus of the endpoints of a vector function $X = X(u, v)$ with three components which is defined on a two-dimensional (u, v)-domain, possesses continuous partial derivatives of n-th order, and is such that the vector product (X_u, X_v) of the first order derivatives does not vanish. If S is a surface of class C^1, then, after a suitable choice of the coordinate axes, S can be represented as the locus of the endpoints of $X = (x, y, z(x, y))$, where $z = z(x, y)$ is a function of class C^1 on some (x, y)-domain. The function $z = z(x, y)$ is of class C^n if and only if the surface S is of class C^n.

The first part of this paper will be concerned with relationships between assumptions of smoothness (that is, degree of differentiability) for a surface S: $X = X(u, v)$ and for its curvatures, either its Gaussian curvature $K = K(u, v)$ or its mean curvature $H = H(u, v)$. These curvatures are defined when S: $X = X(u, v)$ is of class C^n, where $n \geqq 2$, and their definitions show that $H(u, v)$ and $K(u, v)$ are then of class C^{n-2}. Standard theorems in the theory of elliptic differential equations lead to partial converses of this statement.

(i) *If S is a surface having a parametric representation $X = X(u, v)$ of class C^n, where $n \geqq 2$, and if its mean curvature $H(u, v)$ is of class C^n (in u, v), then S is of class C^{n+1} (that is, has some parametric representation of class C^{n+1}).*

If (i) were not a partial converse, but a complete converse, of the facts mentioned above, then the assumption that H is of class C^n could be replaced by the assumption that H is of class C^{n-1}. Such a strengthened form of (i), however, is false; cf. (I**) below.

An analogue of (i), in which H is replaced by K, is true if $K > 0$:

(ii) *If S is a surface having a parametric representation $X = X(u, v)$ of class C^n, where $n > 2$, and if its Gaussian curvature $K(u, v)$ is positive and of class C^n (in u, v), then S is of class C^{n+1}.*

---

* Received March 31, 1952.

This theorem becomes false if the assumption that K is of class C^n is replaced by the assumption that K is of class C^{n-1}; cf. (I**) below. Clearly, (ii) involves two complications not occurring in (i). First, there is in (ii) only the natural restriction on n, namely, $n \geqq 2$, while in (ii) it is assumed that $n > 2$. It will remain an open question whether or not (ii) remains true for $n = 2$. Second, (ii) involves the restriction $K > 0$. Clearly, (ii) is false without this restriction, even in the case $K = \text{const.}$, say $K \equiv 0$ or $K \equiv -1$.

A theorem which meets the first objection raised to (i) and (ii) is as follows:

(I) *Let S be a surface having a parametric representation $S: X = X(u, v)$ of class C^n with the properties that* (α) $n > 2$; (β) *the Gaussian curvature $K(u, v)$ is of class C^{n-1} (in u, v)*; (γ) *the mean curvature $H(u, v)$ is of class C^{n-1} (in u, v)*; (δ) *S has no umbilical points, that is, $H^2 > K$. Then S is of class C^{n+1}.*

It will remain undecided whether or not this theorem is true if (α) is replaced by $n = 2$.

The answer is in the affirmative in the particular case of a constant Gaussian or a constant mean curvature. This follows from (i) for the case of a constant mean curvature. For the case of a Gaussian curvature, the claim is as follows:

(I*) *The assertion of* (I) *remains true if* (α) *and* (β) *are replaced by* (α^*) $n \geqq 2$ *and* (β^*) *the Gaussian curvature $K(u, v) \equiv K_0$ is a constant, respectively.*

While the situation remains problematic for $n = 2$ in the case of an arbitrary $K(u, v)$, the assertion of (I) is of a final nature in the following sense:

(I**) *The assertion* (I) *is false if any one of the assumptions* (β), (γ), (δ) *is omitted (whether $n > 2$ or $n = 2$).*

2. *Remark on a theorem of Scherrer.* The following remark concerns Theorem 2, p. 379, of Scherrer's paper [8]. This theorem states that a surface is determined by the assignment of a first fundamental form, of a mean curvature and of an arbitrary strip.

The interpretation of this statement depends on whether the "determined" (bestimmt) be meant in the sense of the existence of at most one or of at least one surface. In regard to the first interpretation (uniqueness),

it may be worth referring to Bonnet's surfaces (cf. [1], p. 449), those possessing non-trivial isometric deformations which preserve the mean curvature (this class includes all surfaces of constant mean curvature). It would remain to discuss whether two distinct Bonnet surfaces can belong to the same initial strip.

What concerns the second interpretation (existence), the trouble is that there are no general existence theorems for solutions of Scherrer's system of partial differential equations, even when the data are analytic.[1] In order to see this, note that the first fundamental form determines the Gaussian curvature $K(u,v)$. But $H(u,v)$ and $K(u,v)$ must satisfy $H^2 \geqq K$, an inequality which is satisfied on every surface of class C^2 but is not mentioned in the theorem. That the omission of this inequality is not the only trouble is shown by the following counter-example: It is well known (and can be concluded from the formula of Rodrigues) that if a surface has a constant Gaussian curvature, say K_0, and a constant mean curvature, say H_0, then the surface must be part of a sphere, a plane or a circular cylinder according as $H_0{}^2 = K_0 > 0$, $H_0{}^2 = K_0 = 0$ or $H_0{}^2 > K_0 = 0$. In particular, a minimum surface ($H_0 = 0$) cannot have a constant Gaussian curvature $K_0 < 0$. Since the constants $H_0 = 0$, $K_0 = -1$ satisfy the above-mentioned necessary condition $H^2(u,v) \geqq K(u,v)$, it follows that if two given analytic functions H, K of (u,v) satisfy this condition, then there need not exist any surface of class C^2 on which H becomes the mean curvature and K the Gaussian curvature.

3. *Proof of* (i) *and* (ii). Let $X = X(u,v)$ be a parametric representation of class C^n of the surface S, and let $X(u_0, v_0)$ be a given point of S. It can be supposed that the coordinate system $X = (x, y, z)$ has been chosen so that $X(u_0, u_0) = (0, 0, 0)$ and that the unit normal vector at $X(u_0, v_0)$ is $N(u_0, v_0) = (0, 0, 1)$. The last condition implies that the Jacobian $\partial(x,y)/\partial(u,v)$ does not vanish at $(u,v) = (u_0, v_0)$. Hence, S can be represented in the form $z = z(x,y)$, where $z(x,y)$ is of class C^n in a neighborhood of $(x,y) = (0,0)$. If the mean or Gaussian curvature, H or K, is of class C^n in the parameters (u,v), then it is of class C^n in the parameters (x,y) also. In fact, the transformation $(u,v) \rightarrow (x,y)$ is of class C^n with non-vanishing Jacobian, while H (up to a factor ± 1) and K are independent of the choice of parameters.

[1] A corresponding remark seems to hold for Scherrer's Theorem 1 (*loc. cit.*, p. 377). On the other hand, Scherrer's Theorem 3 (*loc. cit.*, p. 380) is surely correct (if $K \neq 0$), but well-known without his apparatus; cf. Weingarten's classical partial differential expression ([1], p. 137) for the sum of the principal curvatures.

Thus there is no loss of generality in assuming in (i) or (ii) that the given parametric representation of S is of the form $X=(x,y,z(x,y))$, where $z(x,y)$ is of class C^n, and $H=H(x,y)$ or $K=K(x,y)$ is of class C^n. In this parametric representation, the function z satisfies the partial differential equations

$$(1)\qquad (1+q^2)r-2pqs+(1+p^2)t-2(1+p^2+q^2)^{3/2}H=0$$

and

$$(2)\qquad rt-s^2-(1+p^2+q^2)^2K=0,$$

where $p=z_x,\ q=z_y,\ r=z_{xx},\ s=z_{xy},\ t=z_{yy}$.

Ad (i). For a given function $H=H(x,y)$, the equation (1) is a partial differential equation for z, say $\phi(x,y,z,p,q,r,s,t)=0$, and is of elliptic type, that is, $4\phi_r\phi_t-\phi_s{}^2>0$. Furthermore, it is linear in the second order derivatives r, s, t, that is, it is of the form $\phi=Ar+Bs+Ct+D$, where A, B, C, D are functions of (x,y,z,p,q). When $H=H(x,y)$ is of class C^n with $n\geqq 2$, these coefficient functions are of class C^n. It follows therefore from considerations of Lichtenstein [5] that any solution $z=z(x,y)$ of class C^2 is of class C^{n+1}; cf. [5], pp. 918-936, in particular, pp. 935-936. (Actually, the assumption that H is of class C^n can be weakened to the assumption that H is of class C^{n-1} and that its derivatives of order $n-1$ satisfy a uniform Hölder condition.)

Ad (ii). For a given positive $K=K(x,y)$, the partial differential equation (2) for z, say $\phi(x,y,z,p,q,r,s,t)=0$, is of elliptic type. When $K=K(x,y)$ is of class C^n with $n>2$, then ϕ is of class C^n in the variables (x,y,z,p,q,r,s,t). It follows therefore from the above-mentioned result of Lichtenstein that if a solution $z=z(x,y)$ of (2) is of class C^n, then it is of class C^{n+1}; cf. [6], pp. 89-90.

The difference between the hypotheses, $n\geqq 2$ and $n>2$, in (i) and (ii) results from the fact that (1) is, but (2) is not, linear in r,s,t. The theorem on a non-linear equation $\phi=0$ is reduced to one dealing with a linear equation by differentiating $\phi=0$ with respect to x (or y) and introducing the new dependent variable $\zeta=z_x$ (or $\zeta=z_y$).

4. *Proof of* (I). As in the last section, it can be supposed that the given parametric representation of S is $X=(x,y,z(x,y))$, where $z(x,y)$ is of class C^n in a neighborhood of $(x,y)=(0,0)$; that $n\geqq 3$; that the curvatures H and K are of class C^{n-1}; and that $H^2>K$. It can also be supposed that

$$z(x,y)=\tfrac{1}{2}(r_0x^2+t_0y^2)+o(x^2+y^2)\ \text{ as }\ (x,y)\to(0,0),$$

that is, that z and the partial derivatives p, q, s vanish at $(x, y) = (0, 0)$; the partial derivatives r, t have the values r_0, t_0 at $(x, y) = (0, 0)$. These normalizations imply that $H(0, 0) = \frac{1}{2}(r_0 + t_0)$ and $K(0, 0) = r_0 t_0$, by (1) and (2) respectively; so that $H^2(0, 0) - K(0, 0) = (r_0 - t_0)^2/4$. Since $(x, y) = (0, 0)$ is not an umbilical point of S, it follows that

$$(3) \qquad r_0 - t_0 \neq 0.$$

Since z is of class C^n, where $n \geqq 3$, the equation (1) can be differentiated with respect to x; the result can be written in the form

$$(4) \qquad (1 + q^2) r_x - 2pqs_x + (1 + p^2) t_x = f_1(x, y),$$

where f_1 is a function of class C^{n-2} (since p, q, H are of class C^{n-1}). Differentiation of (1) with respect to y gives

$$(5) \qquad (1 + q^2) r_y - 2pqs_y + (1 + p^2) t_y = f_2(x, y).$$

Similarly, differentiations of (2) lead to

$$(6) \quad r_x t - 2ss_x + rt_x = f_3(x, y); \qquad (7) \quad r_y t - 2ss_y + rt_y = f_4.$$

As in the case of f_1, the functions f_2, f_3, f_4 are of class C^{n-2}. The derivatives r, s, t satisfy the integrability conditions

$$(8) \qquad r_y - s_x = 0; \qquad (9) \qquad s_y - t_x = 0.$$

The six equations (4)-(9) form a linear system of equations for the six unknowns r_x, s_x, t_x, r_y, s_y, t_y. The matrix of coefficients is

$$\begin{pmatrix} 1 + q^2 & -2pq^2 & 1 + p^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 + q^2 & -2pq & 1 + p^2 \\ t & -2s & r & 0 & 0 & 0 \\ 0 & 0 & 0 & t & -2s & r \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}.$$

It is easily verified that at the point $(x, y) = (0, 0)$, where $p = q = s = 0$ and $r = r_0$, $t = t_0$, the determinant of this matrix is $(r_0 - t_0)^2$, which is not 0 by virtue of (3). By continuity, the determinant does not vanish near $(x, y) = (0, 0)$. Hence, for (x, y) near $(0, 0)$, the equations (4)-(9) can be solved for the six functions $r_x, s_x, \cdots, t_y$ in terms of p, q, r, s, t and f_1, f_2, f_3, f_4. It follows therefore from the formula for the solution of a system of linear equations of non-vanishing determinant that $r_x, s_x, \cdots, t_y$ are of class C^{n-2} (since p, q, r, s, t; f_1, f_2, f_3, f_4 are). But this means that $z(x, y)$ is of class C^{n+1} in a neighborhood of $(0, 0)$. This proves (I).

5. *Proof of* (I*). The proof of (I*) in the three cases $K_0 > 0$, $K_0 < 0$, $K_0 = 0$ is quite different.

Ad $K_0 > 0$. It has recently been shown ([9], p. 366) that if S: $z = z(x, y)$ is of class C^2 and has a constant positive Gaussian curvature, then $z = z(x, y)$ is analytic. This implies (I*) for the case $K_0 > 0$.

Ad $K_0 < 0$. Let S: $z = z(x, y)$ be of class C^2 in a neighborhood of $(x, y) = (0, 0)$, and let S have a constant negative Gaussian curvature K_0, say $K_0 = -1$. Then there exists ([4], pp. 157-161) a transformation $x = x(u, v)$, $y = y(u, v)$ of class C^1, with non-vanishing Jacobian, which leads to a parametric representation $X = X(u, v)$ of S with the properties that $X(u, v)$ is of class C^1, the parametric lines $u =$ const. and $v =$ Const. are asymptotic curves, and the squared line element is of the Tchebychef form,

$$(10) \quad ds^2 = |\, dX \,|^2 = du^2 + 2 \cos \phi \, dudv + dv^2, \text{ where } \phi = \phi(u, v)$$

is continuous and satisfies $0 < \phi(u, v) < \pi$.

Since $X(u, v)$ is of class C^1 only, the standard formulae $h_{ik} = X_{ik} \cdot N$ (where $X_{ik} = \partial^2 X / \partial u^i \partial u^k$, $(u^1, u^2) = (u, v)$, and the period denotes scalar multiplication), defining the elements of the second fundamental matrix, are not applicable. However, since the normal N is of class C^1 with respect to (x, y), it is of class C^1 with respect to (u, v) also. Hence the elements of the second fundamental matrix can be calculated from the formulae $h_{ik} = -X_i \cdot N_k$. (This leads to the same result as the calculation of $h_{ik}(u, v)$ from the second fundamental matrix in the parameters (x, y) by means of the usual transformation rule). The fact that $u =$ const. and $v =$ Const. are asymptotic lines means that $h_{11} \equiv h_{22} \equiv 0$. Since the Gaussian curvature is $\det h_{ik} / \det g_{ik} \equiv -1$, it follows from (10) that $h_{12}{}^2 = \sin^2 \phi$ or $h_{12} = \epsilon \sin \phi$, where $\epsilon = \pm 1$ is independent of (u, v).

Consequently, the mean curvature $H = H(u, v)$ is $-\epsilon \cos \phi / \sin \phi$, where $\sin \phi \neq 0$. The assumption of (I*) concerning H, in the case $n = 2$, is that H is of class C^1 (with respect to (x, y), hence with respect to (u, v)). But this means that $\phi = \phi(u, v)$ is of class C^1. Consequently, Theorem (viii) in [4], p. 163, implies that S: $z = z(x, y)$ is of class C^3. This proves (I*) in the case $K_0 < 0$.

Ad $K_0 = 0$. Let S: $z = z(x, y)$ be of class C^2 in a neighborhood of $(x, y) = (0, 0)$ and let S have the constant Gaussian curvature $K_0 = 0$. On assuming that $(x, y) = (0, 0)$ is not an umbilical point, it can be supposed that

$$(11) \qquad z(x, y) = \tfrac{1}{2} t_0 y^2 + o(x^2 + y^2), \text{ as } (x, y) \to (0, 0),$$

and that

$$t_0 \neq 0. \tag{12}$$

Then

$$x = x, \qquad q = z_y(x, y) \tag{13}$$

defines a transformation $(x, y) \to (x, q)$ of class C^1, with non-vanishing Jacobian $\partial(x, q)/\partial(x, y)$ at $(x, y) = (0, 0)$. (The point $(x, y) = (0, 0)$ corresponds to $(x, q) = (0, 0)$, by (11).)

The inverse transformation $x = x$, $y = y(x, q)$ of (13) is of class C^1 and leads to a parametric representation $X = X(x; q)$ of S, with the properties that $X(x; q)$ is of class C^1 in a vicinity of $(x, q) = (0, 0)$ and that $q = \text{const.}$ is an asymptotic line along which $p = z_x(x, y)$ is a constant, say $p = f(q)$; cf. [3], p. 770. Clearly, $f(q)$ is of class C^1 for small $|q|$.

The asymptotic line, $q = \text{const.}$, is a straight line, which is the intersection of planes having equations of the form

$$f(q)x + qy - z = \alpha(q), \qquad f'(q)x + y = \alpha'(q), \qquad ({}' = d/dq),$$

where $\alpha = \alpha(q)$ is of class C^1 for small $|q|$; cf. [3], p. 770. These equations show that $X = X(x; q)$ is given by

$$\begin{aligned} &x = x, \qquad y = -f'(q)x + \alpha'(q), \\ &z = (f(q) - qf'(q))x - (\alpha(q) - q\alpha'(q)), \end{aligned} \tag{14}$$

for small $x^2 + q^2$; cf. [4], pp. 169-170.

The vector function $X(x, q)$ is of class C^1; so that, since $y = y(x, q) = -f'(q)x + \alpha'(q)$, the functions $\alpha(q)$ and $f(q)$ have continuous second derivatives for small $|q|$.

The relation $p(= z_x) = f(q)$ shows that

$$s = f'(q)t \text{ and } r = f'(q)s = f'^2(q)t. \tag{15}$$

Hence, by (1),

$$\{(1 + q^2)f'^2 - 2pqf' + (1 + p^2)\}t - 2H(1 + p^2 + q^2)^{3/2} = 0.$$

At $(x, y) = (0, 0)$ the coefficient $\{\cdot\ \cdot\ \cdot\}$ of t is $f'^2 + 1 \neq 0$; hence the last equation can be solved for t and shows that t is of class C^1 (since p, q, H and f' are). Thus, by (15), r and s are also of class C^1. Since this means that $z(x, y)$ is of class C^3, the proof of (I*) is complete.

Remark. The derivation of (14) and the arguments in [4], pp. 170-171, imply the last part of the following theorem:

(†) *If S is a surface of class C^2 with Gaussian curvature $K \equiv 0$ and having no flat points, then every sufficiently small piece of S possesses a C^1-parametrization of the form*

$$X(u,v) = A(u)v + B(u), \qquad (A(u) \neq 0), \tag{16}$$

where $A(u)$, $B(u)$ are vector functions of class C^1 satisfying

$$\det(A, A', B') = 0, \qquad ('= d/du). \tag{17}$$

No parametrization of the form (16)-(17) *can be of class C^2 unless S is of class C^3.*

The first part of this assertion is (xiv) in [4], pp. 168-169. The last part is an improvement of (xv), p. 169.

6. *Proof of* (I**). It will be shown that if (α) in (1) is replaced by $n = 2$, then the assertion of (I) becomes false if any one of the remaining conditions, (β), (γ) or (δ), is omitted. It will be clear from the proof that analogous arguments are applicable when condition (α), where $n \neq 2$, is retained.

Omission of (β); $n = 2$. Let $\zeta = \zeta(x,y)$ be of class C^1 on the circle $x^2 + y^2 \leqq 1$ and of class C^2 on the punctured circle $0 < x^2 + y^2 \leqq 1$. Then every solution $z = z(x,y)$ of the Poisson equation

$$r + t = \zeta(x,y) \tag{18}$$

is of class C^2 on $x^2 + y^2 < 1$ and of class C^3 on $0 < x^2 + y^2 < 1$. Let it be granted for a moment that a function $\zeta(x,y)$ can be chosen possessing the smoothness properties specified and having the additional properties that one (hence every) solution z of (18) is not of class C^3 in a neighborhood of $(0,0)$, but that the six third order partial derivatives of z (which exist for (x,y) distinct from, but near, $(0,0)$) satisfy the estimate

$$r_x, s_x, \cdots, t_y = O(|\log(x^2 + y^2)|), \text{ as } (x,y) \to (0,0). \tag{19}$$

Consider the surface S: $z = z(x,y)$, where z is a solution of (18) satisfying

$$p(0,0) = q(0,0) = 0. \tag{20}$$

This surface is of class C^2. Its mean curvature $H = H(x,y)$ is given by

$$2H(1 + p^2 + q^2)^{3/2} = \zeta(x,y) + q^2 r - 2pqs + p^2 t; \tag{21}$$

cf. (1) and (18). Clearly, H is of class C^1 for $0 < x^2 + y^2 < 1$. Actually,

(19) and (20) imply that the first order partial derivatives of H exist, vanish, and are continuous at $(x, y) = (0, 0)$, since p and q are $O(x^2 + y^2)^{\frac{1}{2}}$ as $(x, y) \to (0, 0)$. Hence condition (γ) of (1) is fulfilled.

The function $z = z(x, y)$ can be replaced by $z(x, y) + ax^2 + bxy + cy^2$, without violating (γ). Hence it can be supposed that (δ) is satisfied, that is, that $(x, y) = (0, 0)$ is not an umbilical point (and, therefore, that no point near $(0, 0)$ is umbilical).

Since $z = z(x, y)$ is not of class C^3, it follows that the assertion of (I) can become false when (β) is omitted (in the case $n = 2$), if the existence of a function $\zeta(x, y)$ with the desired properties is granted. The existence of such a function will be verified in Section 7 below.

Omission of (γ); $n = 2$. That (I) is false for $n = 2$ if condition (γ) is omitted, follows either from the fact that there exist torses $(K \equiv 0)$ having no flat point which are of class C^2 but not of class C^3, or from the fact that there exist pseudo-spheres $(K \equiv -1)$ which are of class C^2 but not of class C^3 (cf. [2], Part 1).

Omission of (δ); $n = 2$. Let $\zeta = \zeta(x, y)$ be the function occurring above in connection with (18). Let $z = z(x, y)$ be a solution of (18) satisfying (20). Finally, let the constants a, b, c be chosen so that $Z(x, y) = z(x, y) + ax^2 + bxy + cy^2$ satisfies

$$(22) \qquad Z_{xx} = Z_{xy} = Z_{yy} = 0 \text{ at } (x, y) = (0, 0).$$

The surface S: $z = Z(x, y)$ is of class C^2 but not of class C^3. By the argument following (21), the mean curvature $H = H(x, y)$ of S is of class C^1; so that (γ) holds. The Gaussian curvatures $K = K(x, y)$ is of class C^1 for $0 < x^2 + y^2 < 1$. It also follows from (2) and (19), where p, q, r, s, t represent the partial derivatives of Z, that the first order partial derivatives of K exist, vanish, and are continuous at $(x, y) = (0, 0)$ (in fact, (19) and (22) imply that r, s, t are $O((x^2 + y^2)^{\frac{1}{2}} \,|\log(x^2 + y^2)|)$ as $x^2 + y^2 \to 0$).

This proves that when condition (δ) is omitted, (I) becomes false (in the case $n = 2$), if the existence of a function $\zeta(x, y)$ with the desired properties is granted.

It may be mentioned that this example can be modified so as to make K positive. Such an example results by putting $z = Z(x, y) + \frac{1}{2}(x^2 + y^2)$. The mean curvature H of this surface is of class C^1 (cf. the paragraph following (21)). The Gaussian curvature is $1 + Z_{xx} + Z_{yy} + K(x, y)$, where K is the Gaussian curvature of $z = Z$. Since $Z_{xx} + Z_{yy}$ is of class C^1, the Gaussian curvature of $z = Z + \frac{1}{2}(x^2 + y^2)$ is of class C^1.

7. *The existence of* $\zeta(x, y)$. The proof of (I**) will be complete if it is shown that there exists a function $\zeta(x, y)$ defined on the circle $x^2 + y^2 \leqq 1$ such that the solutions of the Poisson equation (18) have the properties enumerated after (18). To this end, first consider Petrini's example ([7], p. 138) of a Poisson equation,

$$u_{xx} + u_{yy} = 2\pi\omega(x, y) \tag{23}$$

with $\omega(0, 0) = 0$ and

$$\omega(x, y) = x^2/\{(x^2 + y^2)\,|\log(x^2 + y^2)|\} \text{ if } x^2 + y^2 \neq 0,$$

which equation has no solution (of class C^2) although ω is continuous. Actually, the logarithmic potential

$$u(x, y) = \iint\limits_{\xi^2+\eta^2 \leqq 1} \omega(\xi, \eta) \log\{(\xi - x)^2 + (\eta - y)^2\}^{\frac{1}{2}} d\xi d\eta \tag{24}$$

is of class C^1 on $x^2 + y^2 < 1$, of class C^2 (in fact, analytic) on $0 < x^2 + y^2 < 1$, and satisfies (23) there, but neither of the partial derivatives u_{xx}, u_{yy} exists at $(x, y) = (0, 0)$.

It should be mentioned, for later reference, that

$$u_x(0, y) = 0 \text{ for } -1 < y < 1, \tag{25}$$

since (24) is an even function of x.

It will be shown that the second order partial derivatives of (24) satisfy

$$u_{xx}, u_{xy}, u_{yy} = O(|\log(x^2 + y^2)|) \text{ as } (x, y) \to (0, 0). \tag{26}$$

If $(x, y) \neq (0, 0)$ and if $R = R(x, y) > 0$ is sufficiently small and fixed, then $u_{xx}(x, y)$ is the sum of $I_0 = \pi\omega(x, y)$,

$$I_1 = \lim_{h \to 0} \int_0^{2\pi} \cos 2\theta \, \{\int_h^R \omega(x + \rho\cos\theta, y + \rho\sin\theta)\rho^{-1} d\rho\} d\theta$$

and

$$I_2 = \iint\limits_D \omega(\xi, \eta)(\eta - y)^2\{(\xi - x)^2 + (\eta - y)^2\}^{-2} d\xi d\eta,$$

where D is the region bounded by $\xi^2 + \eta^2 = 1$ and $(\xi - x)^2 + (\eta - y)^2 = R^2$; cf. [7], pp. 132-134.

Clearly, $|I_0| = O(1)$ as $(x, y) \to (0, 0)$. Let R be fixed and let $x^2 + y^2 = (2R)^2$. Since ω is bounded,

$$I_2 = O\left(\iint_D (\sin^2\theta)(\rho^{-2})\rho d\rho d\theta\right) = O(|\log R|),$$

so that $I_2 = O(|\log(x^2 + y^2)|)$. In order to appraise I_1, note that the value of I_1 is not affected if $\omega(x + \rho\cos\theta, y + \rho\sin\theta)$ is replaced by $\omega(x + \rho\cos\theta, y + \rho\sin\theta) - \omega(x, y)$. For $0 < h \leqq R$, $0 \leqq \theta \leqq 2\pi$, the latter difference is majorized by const. ρ times the maximum of

$$|\partial\omega(x, y)/\partial x| + |\partial\omega(x, y)/\partial y| \quad \text{for } 1 \geqq x + y \geqq R^2.$$

Hence, for $x^2 + y^2 = (2R)^2$, $0 < h \leqq \rho \leqq R$, the definition of ω implies that

$$|\omega(x + \rho\cos\theta, y + \rho\sin\theta) - \omega(x, y)| \leqq O(\rho R^{-1}|\log R|).$$

Consequently, $I_1 = O(|\log R|) = O(|\log(x^2 + y^2)|)$. This proves the estimate (26) for u_{xx}. The estimates for u_{xy} and u_{yy} are proved similarly.

Define $\zeta(x, y)$ by the relation

$$\zeta(x, y) = 2\pi \int_0^x \omega(t, y)\, dt. \tag{27}$$

It will be shown that $\zeta(x, y)$ is a function having the properties specified in connection with (18). First, $\zeta(x, y)$ is of class C^1 for all (x, y). In fact,

$\zeta_x = 2\pi\omega$, while $\zeta_y = 2\pi \int_0^x \omega_y(t, y)dt$, and it is readily verified from the definition of ω that the last integral exists (as a Riemann integral) and is a continuous function of (x, y), even at $(x, y) = (0, 0)$. It is also clear from (27) that $\zeta(x, y)$ is of class C^2 (in fact, analytic) for $x^2 + y^2 > 0$.

All solutions $z = z(x, y)$ of the Poisson equation (18) are, therefore, of class C^2 on $x^2 + y^2 < 1$ and of class C^3 on $0 < x^2 + y^2 < 1$. It remains to show that $z(x, y)$ is not of class C^3 in a neighborhood of $(x, y) = (0, 0)$, and that the third order partial derivatives of z satisfy (19). Since the solutions of (18) differ only in (additive) harmonic functions, it is sufficient to verify this for a particular solution of (18).

To this end, put

$$z(x, y) = \int_0^x u(t, y)\, dt, \tag{28}$$

where $u(x, y)$ is defined by (24). The function (28) has the first order partial derivatives $z_x = u$, $z_y = \int_0^x u_y(t, y)\,dt$, since $u(x, y)$ is of class C^1. It also has the second order partial derivatives

$$(29)\qquad \begin{aligned} &z_{xx} = u_x, \qquad z_{xy} = z_{yx} = u_y \text{ for } x^2 + y^2 < 1,\\ &z_{yy} = \int_0^x u_{yy}(t, y)\,dt \text{ for } x^2 + y^2 < 1, y \neq 0. \end{aligned}$$

Since u is of class C^2 and satisfies (22) for $0 < x^2 + y^2 < 1$, it follows that $z_{yy} = 2\pi \int_0^x \omega(t, y)\,dt - \int_0^x u_{xx}(x, y)\,dt$, hence $z_{yy}(x, y) = \zeta(x, y) - u_x(x, y) + u_x(0, y)$, for $x^2 + y^2 < 1$ and $y \neq 0$. By (25), this becomes

$$(30)\qquad z_{yy}(x, y) = \zeta(x, y) - u_x(x, y) \text{ for } x^2 + y^2 < 1 \text{ and } y \neq 0.$$

In view of the continuity of the right side of (30) (even for $y = 0$), the function z has for $x^2 + y^2 < 1$ the continuous second partial derivative z_{yy}, given by (30). Thus, by (29) and (30), the function (28) is of class C^2 for $x^2 + y^2 < 1$ and satisfies (18).

Since u_{xx} and u_{yy} fail to exist at $(x, y) = (0, 0)$, the derivatives z_{xxx}, z_{xyy} do not exist at $(x, y) = (0, 0)$. Hence $z(x, y)$ is not of class C^3 near $(x, y) = (0, 0)$. However, it is of class C^3 (even analytic) for $0 < x^2 + y^2 < 1$, since it is a solution of (18).

Finally, the third order partial derivatives of z, that is,

$$r_x = u_{xx},\ s_x = u_{xy},\ t_x = u_{yy},\ r_y = u_{xy},\ s_y = u_{yy} \text{ and } t_y = \zeta_y - u_{xy},$$

satisfy (19) by virtue of (26). This completes the proof of the existence of the function $\zeta(x, y)$ granted in the last section.

8. *Smoothness of the second fundamental form.* It is indicated by (I) that, under certain general assumptions, an unexpected smoothness of the coefficients h_{ik} of the second fundamental form

$$(31)\qquad -dX \cdot dN = h_{ik}du^i du^k, \text{ where } (u^1, u^2) = (u, v),$$

cannot occur. The following theorem is in the same direction.

(II) *If S is a surface having a parametric representation $X = X(u, v)$ of*

class C^n, *for some* $n \geqq 2$, *with the property that the coefficients* $h_{ik} = h_{ik}(u, v)$ *in* (31) *are of class* C^{n-1}, *then* S *is of class* C^{n+1}.

This theorem is implied by (I) only if $n > 2$ and if S has no umbilical points.

Remark. It will be clear from the proof of Corollary 1 below that (II) *remains true if only the two functions* h_{11}, h_{22} (rather than, as in the above wording, all three functions $h_{11}, h_{12} = h_{21}, h_{22}$) *are assumed to be of class* C^{n-1}.

COROLLARY 1. *If* S *is a surface having a parametric representation* $X = X(u, v)$ *of class* C^n, $n \geqq 2$, *for which* $h_{11} = h_{22} = 0$ (*i. e.*, $u =$ const. *and* $v =$ Const. *are asymptotic lines*), *then* S *is of class* C^{n+1}.

This is known if the Gaussian curvature K is a negative constant ([4], Theorem (viii), p. 163), or more generally when $K(u, v)$ (< 0) is of class C^1 ([2], p. 223). It completes Theorem (v) in [4], p. 156.

The assertion of Corollary 1 becomes false if the asymptotic lines of S are replaced by the lines of curvature. In order to see this, it is sufficient to consider a surface of revolution in terms of its parameters (u, v), where u is the arc length on a meridian $v =$ const. and the latter is a curve which is of class C^n but not of class C^{n+1}. Such a surface of revolution is of class C^n in (u, v), has no parametrization of class C^{n+1}, and $u =$ const., $v =$ Const. are its lines of curvature, if it has no umbilical points.

What corresponds to Corollary 1 in case of lines of curvature is contained in the following assertion:

COROLLARY 2. *If* S *is a surface having a parametric representation* $X = X(u, v)$ *of class* C^n, $n \geqq 2$, *for which* $h_{12} \equiv 0$ (*e. g.*, $u =$ const. *and* $v =$ Const. *are lines of curvature*) *and for which either the mean curvature* $H(u, v)$ *is of class* C^{n-1} *or the Gaussian curvature* $K(u, v)$ *is of class* C^{n-1} *and does not vanish, then* S *is of class* C^{n+1}.

The proof of (II) has been given for a particular case in [4], p. 163.

Proof of (II). Since $X(u, v)$ is of class C^n, where $n \geqq 2$, the Weingarten derivation formulae for the partial derivatives $N_1 = N_u$, $N_2 = N_v$ of the unit normal vector are valid,

$$N_i = -g^{jk} h_{ij} X_k, \tag{32}$$

where (g^{jk}) is the inverse matrix of the matrix of coefficients of the first fundamental form

$$dX \cdot dX = g_{ik} du^i du^k. \tag{33}$$

It is clear from (33) that the functions $g_{ik} = g_{ik}(u, v)$ are of class C^{n-1} (since X is of class C^n). Hence, by (32), N_1 and N_2 are of class C^{n-1}; that is, $N = N(u, v)$ is of class C^n.

Let (u_0, v_0) be a point of S, and suppose that $N(u_0, v_0) = (0, 0, 1)$; so that $\partial(x, y)/\partial(u, v) \neq 0$ at (u_0, v_0), if $(x, y, z) = X = X(u, v)$. Thus the transformation $(u, v) \to (x, y)$ is of class C^n with non-vanishing Jacobian at (u_0, v_0). The inverse transformation leads to the Cartesian form $z = z(x, y)$ for S. Hence $z(x, y)$ is of class C^n.

Since $N(u, v)$ is of class C^n and, up to a factor ± 1, is independent of the parameters, it follows that N is of class C^n as a function of (x, y). But $N = \pm(z_x, z_y, -1)/(1 + z_x{}^2 + z_y{}^2)^{\frac{1}{2}}$; consequently, z_x and z_y are of class C^n, which means that z is of class C^{n+1}. This proves (II).

The proofs of Corollaries 1 and 2 will depend on the circumstance that the equations of Codazzi, in an integrated form, are valid on surfaces of class C^2 ([3], pp. 759-760). These equations are

$$\int_J h_{ik} du^k = \iint_D I_i \, dudv \qquad (i = 1, 2,), \tag{34}$$

where it is assumed that $X(u, v)$ is of class C^2 on a simply connected (u, v)-domain, J is a Jordan curve which is piece-wise of class C^1, D is the interior of J, I_i denotes $\Gamma^k{}_{i1} h_{k2} - \Gamma^k{}_{i2} h_{k1}$, and $\Gamma^i{}_{jk} = \Gamma^i{}_{jk}(u, v)$ are the Christoffel symbols of the second kind. The latter are of class C^{n-2} if X is of class C^n.

Proof of Corollary 1. When $h_{11} = h_{22} = 0$, the equations (34) become

$$\int_J h_{12} dv = \iint_D I_1 dudv, \tag{35}$$

$$\int_J h_{12} du = \iint_D I_2 dudv \tag{36}$$

respectively. Let J be chosen to be the positively oriented rectangle with the vertices (u, v), $(u + \Delta u, v)$, $(u + \Delta u, v + \Delta v)$, $(u, v + \Delta v)$. On dividing (35) by $\Delta v \to 0$, it is seen that

$$h_{12}(u + \Delta u, v) - h_{12}(u, v) = \int_u^{u+\Delta u} I_1 \, du.$$

It follows that $\partial h_{12}/\partial u$ exists and is of class C^{n-2}. Similarly, it follows from (36) that $\partial h_{12}/\partial v$ exists and is of class C^{n-2}. Hence h_{12} is of class C^{n-1}, and Corollary 1 follows from (II).

Proof of Corollary 2. Arguing as in the last proof, it is seen that when $h_{12} \equiv 0$, the equations (34) imply that $\partial h_{11}/\partial v$, $\partial h_{22}/\partial u$ exist and are of class C^{n-2}.

On the other hand, the definitions of H and K show that

$$g_{22}h_{11} + g_{11}h_{22} = 2H(g_{11}g_{22} - g_{12}^2) \text{ and } h_{11}h_{22} = K(g_{11}g_{22} - g_{12}^2).$$

If H is of class C^{n-1}, it follows from $g_{22} \neq 0$ that $\partial h_{11}/\partial u$ exists and is of class C^{n-2}. Similarly, if K is not 0 and is of class C^{n-1}, then h_{22} is not 0 and $\partial h_{11}/\partial u$ exists and is of class C^{n-2}. In either case, h_{11} is of class C^{n-1}. Thus h_{22} is of class C^{n-1} and Corollary 2 follows from (II).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] L. Bianchi, *Lezioni di geometria differenziale*, Pisa, 1894.

[2] P. Hartman, "On unsmooth two-dimensional Riemannian metrics," *American Journal of Mathematics*, vol. 74 (1952), pp. 215-226.

[3] ——— and A. Wintner, "On the fundamental equations of differential geometry," *ibid.*, vol. 72 (1950), pp. 757-774.

[4] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[5] L. Lichtenstein, "Ueber den analytischen Charakter der Lösungen regulärer zweidimensionaler Variationsprobleme," *Bulletin de l'Académie des Sciences de Cracovie*, ser. A, 1913, pp. 915-941.

[6] ———, *Vorlesungen über einige Klassen nicht-linearer Integralgleichungen und Integro-Differentialgleichungen*, Berlin, 1931.

[7] H. Petrini, "Les dérivées premières et secondes du potential logarithmique," *Journal de Mathématique*, ser. 6, vol. 5 (1909), pp. 127-223.

[8] W. Scherrer, Stützfunktion und Radius. I., *Commentarii Mathematici Helvetici*, vol. 20 (1947), pp. 366-381.

[9] A. Wintner, "On parallel surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 365-376.

ENVELOPES AND DISCRIMINANT CURVES.*

By PHILIP HARTMAN and AUREL WINTNER.

1. Moigno's edition of Cauchy's lectures contains a passage concerning the possible existence of an envelope for the solution curves of a real differential equation

$$(1) \qquad y' = f(x, y),$$

where f is *single-valued* and continuous. The passage in question ([1], pp. 445-451) is quite obscure; both the formulation and the proof of the criterion are far from the standards of Cauchy's customary precision. This may be a reason why Cauchy's criterion seems to have been neglected. Actually, it can easily be put into a modern form, which is the content of (i) below.

Another reason for the neglect afforded to Cauchy's envelope criterion seems to be the circumstance that his remarks refer to a "solved" differential equation (1), whereas it is the "unsolved" form,

$$(2) \qquad F(x, y; y') = 0,$$

which usually leads to envelopes in the applications; for instance, if F is a polynomial (of degree $n > 1$) in y'. Actually, the latter situation can be reduced to the former by an application of the following principle:

Suppose that, when (x, y) is confined to a sufficiently small domain, the algebraic equation (2), the coefficients of which are given continuous functions of (x, y), has a certain number of *real* roots, say $y' = f_k(x, y)$, where $k = 1, \cdots, m$ (so that $m \leqq n$). Then, within that (x, y)-domain, (2) is formally equivalent to the system of m "solved" equations, those which result by choosing $f = f_k$ in (1); so that (2) *can be thought of as a superimposition of m "velocity fields"* (1). Hence, if criterion (i), which deals with (1), assures the existence of envelopes for one or more of the resulting m equations (1), then there result envelopes for (2) also. [On the other hand, the converse inference requires caution; in this regard, cf. Perron's results [5] on "solving" (2) with respect to y'.]

Theorem (ii) below will illustrate this principle by applying (i) to the classical problem of nets (in differential geometry), where F is a quadratic

* Received May 31, 1952.

polynomial in y' or, more symmetrically as well as more generally, (2) is given by a quadratic form in (dx, dy), say

$$(3) \qquad a(x,y)\,dx^2 + 2b(x,y)\,dx\,dy + c(x,y)\,dy^2 = 0.$$

Here a, b, c are continuous functions which, in order that (3) has (real) solution curves, must have a non-positive discriminant

$$(4) \qquad D(x,y) \equiv D = ac - b^2$$

in the (x, y)-region under considerations. Under suitable assumptions of smoothness (for instance, if the coefficients of (3) satisfy a uniform Lipschitz condition), the envelopes, if any, of the solution paths of (3) will have to be found amongst the "branches" of the "discriminant curve"

$$(5) \qquad D(x,y) = 0.$$

(The converse is not in general true, since, even in the analytic case, a (real) branch of (5) need not be an envelope.)

It is clear from the remarks made before (i) that the *envelope* situations delimited by (i), and therefore by (ii), are considered as resulting from points of non-uniqueness of an ordinary differential equation (1). Correspondingly, an opposite situation possible on a branch of the discriminant curve, that corresponding to *tac-points,* can be thought of as representing points of uniqueness. This is the actual content, as well as the proof, of (iii) below.

As an illustration of (ii), consider the case in which (3) represents the equation of the asymptotic curves on a (sufficiently smooth) surface $\mathfrak{S}$: $z = z(x,y)$ of non-negative Gaussian curvature $K = K(x,y)$, and suppose that $K \neq 0$ except on a certain parabolic curve, $\mathfrak{S}^0$: $K(x,y) = 0$, on $\mathfrak{S}$. In Section **7**, conditions will be obtained under which $\mathfrak{S}^0$ is an envelope of the asymptotic curves contained in the hyperbolic region of $\mathfrak{S}$.

Theorems (iv) and (v) will deal with (a branch of) a discriminant curve which is "a singular line" of (3), in the sense that all three coefficients of (3) vanish at every point of the curve (4).

As an illustration of the results for the case of such a singular line of (3), the behavior of the lines of curvature near a line of umbilical points on a surface will be discussed in Section **10**.

2. For small non-negative y, say for $0 \leqq y \leqq \beta$, let ϕ be a continuous function satisfying the three conditions

$$(6) \qquad \phi(y) > 0 \text{ if } y > 0, \qquad \phi(0) = 0, \qquad \int_{+0} dy/\phi(y) < \infty,$$

and let $x = x(y)$ denote the value of the last integral when the upper limit of integration is y. Then, since $x(y)$, where $x(0) = 0$, is a strictly increasing continuous function, there exists a unique inverse function, $y = y(x)$, satisfying $y(x) \gtreqqless 0$ according as $x \gtreqqless 0$, where $x \leqq \alpha$ if α denotes $x(\beta)$. It is readily verified that, if c is an arbitrary constant, then $y = y(x - c)$, where $c \leqq x \leqq c + \alpha$, is a solution of the case $f(x, y) = \phi(y)$ of (1). Since this solution of $y' = \phi(y)$ has at $x = c$ the (unilateral) slope $\phi(0) = 0$, and since $y(x) \equiv 0$ is another solution, it follows that the x-axis is an envelope of solutions.

This situation can be transferred from $y' = \phi(y)$ to a more general equation (1), as follows:

(i) *On a rectangle* $\mathfrak{R}$ *of the form*

$$\mathfrak{R}: -a \leqq x \leqq a, \; 0 \leqq y \leqq b, \tag{7}$$

let $f(x, y)$ *be a continuous function satisfying*

$$f(x, 0) \equiv 0 \qquad (-a \leqq x \leqq a) \tag{8}$$

and having the property that, for some continuous function $\phi(y)$ *satisfying* (6),

$$f(x, y) > \phi(y) \text{ if } -a \leqq x \leqq a, \; 0 < y \leqq b. \tag{9}$$

Put $M = \max_{\mathfrak{R}} f(x, y)$ *and let* $\mathfrak{Q}$ *denote the* (x, c)*-set*

$$\mathfrak{Q}: c \leqq x \leqq \min(a, c + b/M), \; -a < c < a \tag{10}$$

in an (x, c)*-plane. Then there exists on* $\mathfrak{Q}$ *a function* $y = y(x; c)$ *which is continuous from the left with respect to* c *(uniformly in* x*) and has the following properties:* $y(x; c)$ *is a solution of* (1) *(for fixed* c*), and* $y(x; c) \gtreqqless 0$ *holds according as* $x \gtreqqless c$*; finally (a portion of) the* x*-axis is an envelope of this one-parameter family of solutions.*

Although $y(x; c)$ is continuous from the left in c (uniformly in x) and has a (continuous) derivative with respect to x, the continuity of $y(x; c)$ in c cannot in general be asserted.

Proof of (i). Let $y_0(x)$, where $0 \leqq x \leqq \min(a, b/M)$, denote the function which was denoted by $y(x)$ in the remarks made after (6). Then, if x and c are subject to the inequalities (10), the function $z = y_0(x - c)$ is a solution of $z' = \phi(z)$ and satisfies $y_0(x - c) \gtreqqless 0$ according as $x \gtreqqless c$. In addition, $z = y_0(x - c)$ is the maximal solution (in the sense of Osgood [4]), with reference to the initial condition $z(c) = 0$ and to $x \geqq c$, of the

differential equation $z' = \phi(z)$; cf. [6]. Hence, if $y(x; c)$ denotes, within the restrictions (10), that solution $y(x)$ of (1) which is maximal with reference to the initial condition $y(c) = 0$ and to $x \geqq c$, it is clear from (9) and from the first two of the three assumptions (6) that $y(x; c) > y(x - c)$ if $x > c$. This, when compared with the assumption (8), proves the assertions of (i) except the continuity assertion, $y(x; c) \to y(x; c_0)$ as $c \to c_0 - 0$ (uniformly in x). But the latter assertion merely expresses a well-known property of Osgood's maximal solutions.

Remark. It is clear from the above proof of (i) that assumption (9) can be generalized to the assumption of minorants ϕ which depend not only on y but also on x and c and have the property that, corresponding to the assumption (6) for $z' = \phi(z)$, the non-uniqueness of the initial value problem $z(c) = 0$ of $z' = \phi(x, z; c)$ is assured (for small $x - c \geqq 0$).

3. Assumption (8) of (i), rendering the x-axis an envelope, is just a normalization; by a change of variables, it can be replaced by

$$y^{0\prime}(x) - f(x, y^0(x)) \equiv 0, \tag{8 bis}$$

if the solution curve $y = y^0(x)$, instead of $y = y(x) \equiv 0$, is suspected to be an envelope. Such a change of variables will be needed in the proof of the following illustration of the principle italicized in Section **1**.

On a (sufficiently small) domain of the (x, y)-plane, say on the circle $\mathfrak{C}_a$: $x^2 + y^2 < a^2$, let the three coefficients of (3) be functions of class C^1 satisfying

$$D(0,0) = 0 \quad \text{and} \quad \operatorname{grad} D(0,0) \neq 0, \tag{11}$$

where $D = D(x, y)$ is the discriminant (4), and denote by $\mathfrak{D}^+{}_a$, $\mathfrak{D}^- = \mathfrak{D}^-{}_a$ and $\mathfrak{D}^0 = \mathfrak{D}^0{}_a$ the sets of those points (x, y) of $\mathfrak{C} = \mathfrak{C}_a$ which satisfy $D(x, y) > 0$, $D(x, y) < 0$ and (5), respectively. It is clear from (11) and from the (local) existence theorem of implicit functions that, if the radius a is small enough, $\mathfrak{D}^+$ and $\mathfrak{D}^-$ are non-empty open sets separated by the discriminant curve $\mathfrak{D}^0$ which is a Jordan arc of class C^1 passing through the point $(0, 0)$.

It is also clear from (4) and (11) that, for reasons of reality, no point of $\mathfrak{D}^+$ issues any solution path of (3), while every point of $\mathfrak{D}^-$ issues solution paths in two distinct directions. Thus (3) splits near every point of $\mathfrak{D}^-$ into two distinct equations of the type (1) (with x and y possibly interchanged), leading to the following situation: All those solution paths of (3) which have no point on the discriminant curve fall into two families, say

$\mathfrak{F}_1$ and $\mathfrak{F}_2$, in such a way that each $\mathfrak{F}_j$ affords a *schlicht* covering of $\mathfrak{D}^-$ and no curve of $\mathfrak{F}_1$ touches any curve of $\mathfrak{F}_2$.

Nothing follows in this manner with regard to the behavior of either $\mathfrak{F}_j$ at the boundary $\mathfrak{D}^0$ of the open set $\mathfrak{D}^-$. The latter question can however be discussed with the help of the above criterion (i).

Let the ratio $\mu:\lambda$ or $-\mu:\lambda$, where $\mu=0$ is allowed but $\lambda=0=\mu$ is not, be the slope of $\mathfrak{D}^0$ at $(0,0)$. Then two cases are possible, according as this slope does or does not have the same value as the slope $dy_0:dx_0$ or $-dy_0:dx_0$ which, in view of (4) and (11), is determined by (3) at $(0,0)$ (so that the two cases are characterized by

$$(12)\qquad a(0,0)\lambda^2+2b(0,0)\lambda\mu+c(0,0)\mu^2=0$$

and

$$(12^*)\qquad a(0,0)\lambda^2+2b(0,0)\lambda\mu+c(0,0)\mu^2\neq 0$$

respectively).

(ii) *Let the coefficient functions of* (3), *hence also its discriminant* (4), *be functions of class C^1 on a domain, say on $\mathfrak{C}_a: x^2+y^2<a^2$, where a is sufficiently small, and suppose that $D(x,y)$ satisfies* (11) *at* $(0,0)$. *Suppose further that the discriminant curve $\mathfrak{D}^0$ (which then exists and is of class C^1) is a solution path of* (3). *Then $\mathfrak{D}^0$ is an envelope of $\mathfrak{F}_j$, and every point of $\mathfrak{D}^0$ issues (in the direction of $\mathfrak{D}^0$) at least one curve contained in $\mathfrak{F}_j$, where $j=1,2$.*

Under the assumptions stated, a point of the envelope *can* issue more than two branches contained in the net $\mathfrak{F}_1+\mathfrak{F}_2$.

Proof of (ii). According to (4) and (11), the matrix of (3) at $(0,0)$ has the determinant zero but is not the zero matrix and possesses therefore exactly one non-vanishing eigenvalue. Hence, after a rotation of the (x,y)-plane, it can be assumed that

$$(13)\qquad a(0,0)=0,\qquad b(0,0)=0,\qquad c(0,0)\neq 0.$$

Since $\mathfrak{D}^0$ is supposed to be a solution path of (3), its slope $\pm\mu:\lambda$ at $(0,0)$ must satisfy (12). In view of (13), this means that $\mu=0$, i. e., that the x-axis is tangent to $\mathfrak{D}^0$ at $(0,0)$. It follows that $\mathfrak{D}^0$ is of the form

$$(14)\qquad \mathfrak{D}^0:\ y=y^0(x);\qquad y^0(0)=0,\qquad y^{0\prime}(0)=0$$

(provided that $|x|$ is small enough).

It also follows from (13) that

$$(15)\qquad c(x,y)\neq 0\quad \text{on}\quad \mathfrak{C}_a:\ x^2+y^2<a^2,$$

if a is small enough. Hence (3) on $\mathfrak{C}_a = \mathfrak{D}^+ + \mathfrak{D}^- + \mathfrak{D}^0$, being equivalent to (3) on $\mathfrak{D}^- + \mathfrak{D}^0$, is equivalent to the two differential equations

$$(16_j) \qquad y' = f_j(x, y) \equiv (-b + (-1)^j | D |^{\frac{1}{2}})/c$$

on $\mathfrak{D}^- + \mathfrak{D}^0$, where $j = 1, 2$, and both functions f_j are real-valued and continuous on $\mathfrak{D}^- + \mathfrak{D}^0$. In addition, f_j is of class C^1 on $\mathfrak{D}^-$, and (16_j) defines on $\mathfrak{D}^-$ the family which, before (ii) above, was denoted by $\mathcal{F}_j$. Note that $f_1(x, y) = f_2(x, y)$ if and only if (x, y) is a point of the discriminant curve (14).

4. Since $\mathfrak{D}^0$ is a solution path of (3), it follows from (14) and either of the equations (16_j) that

$$(17) \qquad y^{0\prime}(x) \equiv -b(x, y^0(x))/c(x, y^0(x))$$

is an identity in x. If this identity is subtracted from (16_j) and if $y - y_0(x)$ is then denoted simply by y, it follows that (16_j) is equivalent to the case

$$(18_j) \qquad f(x, y) = f_j(x, y + y^0(x)) - f_j(x, y^0(x))$$

of (1). Both the functions (18_j) satisfy assumption (8) of (i), since (14) becomes $y(x) \equiv 0$ in the present notations. Hence, in order to prove (ii), it is sufficient to ascertain that both functions (18_j) satisfy the remaining assumptions of (i).

Since $\mathfrak{D}^0$ is now a segment of the x-axis, $\mathfrak{D}^-$ is either the portion $y > 0$ or the portion $y < 0$ of the circle $\mathfrak{C}_a$: $x^2 + y^2 < a^2$. Clearly, the latter case can be reduced to the former. Hence it can be assumed that $\mathfrak{D}^- + \mathfrak{D}^0$ contains a rectangle of the form (7). On the other hand, assumption (6) of (i) is satisfied if $\phi(y)$ is a positive constant multiple of $y^{\frac{1}{2}}$, and it is clear from the proof of (i) that assumption (9) can be replaced by $f(x, y) < -\phi(y)$. Consequently, the proof of (ii) will be complete if it is ascertained that there belongs to the functions (18_1), (18_2) a positive constant satisfying

$$(19_1) \qquad f(x, y) \geqq \text{Const. } y^{\frac{1}{2}}; \qquad\qquad (19_2) \qquad f(x, y) \leqq -\text{Const. } y^{\frac{1}{2}}$$

at every point of a sufficiently small rectangle (7).

First, if $g(x, y)$ denotes the function

$$(20) \quad g(x, y) = b(x, y + y^0(x))/c(x, y + y^0(x)) - b(x, y^0(x))/c(x, y^0(x)),$$

then, since $b(x, y)$, $c(x, y)$ and $y^0(x)$ are functions of class C^1, it is clear from the above normalizations that $| g(x, y) |$ is majorized by a constant multiple of y (on a sufficiently small rectangle (7)). On the other hand,

it is seen from (20) and (16_j), where $|D| = -D$, that (18_j) can be written in the form

$$(18_j\,\mathrm{bis}) \qquad -f(x,y) = g(x,y) - (-1)^j(-D)^{\frac{1}{2}}/c,$$

where the argument of $(-D)^{\frac{1}{2}}/c$ is $(x, y + y^0(x))$.

Hence, in order to prove the existence of a positive constant with reference to which (18_j) satisfies (19_j), it is sufficient to show that the value of $|(-D)^{\frac{1}{2}}/c|$ at $(x, y + y^0(x))$ is minorized by a positive constant multiple of $y^{\frac{1}{2}}$. It follows therefore from (15) that it is sufficient to prove the existence of a positive constant satisfying $|D(x, y + y^0(x))| \geqq \mathrm{const.}\, y$ at every point of a sufficiently small rectangle (7). Hence it is seen from the last two of the relations (14) and from Taylor's formula that it is sufficient to prove the non-vanishing of the partial derivative $D_y(0,0)$. It follows therefore from the second of the assumption (11) that it is sufficient to prove the vanishing of $D_x(0,0)$.

Finally, $D_x(0,0) = 0$ is equivalent to the statement that the slope of the discriminant curve $\mathfrak{D}^0$ vanishes at $(x,y) = (0,0)$. Since the truth of this statement is implied by the above normalization of $\mathfrak{D}^0$, the proof of (ii) is now complete.

5. The last assumption of (ii), according to which the discriminant curve is a solution path, has implied (12). In what follows, the case (12*) will be considered.

(iii) *Suppose that all but the last of the assumptions of* (ii) *are satisfied but that the discriminant curve* $\mathfrak{D}^0$, *instead of being a solution path of* (3), *violates not only* (17) *but even the case* $x = 0$ *of* (17), *which is* (12); *so that* (12*) *is satisfied. Then, if the arc* $\mathfrak{D}^0$ *is chosen short enough, every* (x^0, y^0) *on* $\mathfrak{D}^0$ *issues exactly one solution path of the family* $\mathfrak{F}_j$, *where* $j = 1, 2$ (*and neither of these solution paths is tangent to* $\mathfrak{D}^0$; *so that there is no envelope*).

Proof of (iii). For reasons of continuity, it can be assumed that (x^0, y^0) is the point $(0,0)$ and, after a rotation of the (x,y)-plane, that the tangent of $\mathfrak{D}^0$ at $(0,0)$ is the y-axis. This means that $\mathfrak{D}^0$ is of the form

$$(21) \qquad \mathfrak{D}^0\colon\ x = x^0(y); \qquad\qquad x^0(0) = 0, \quad x_y{}^0(0) = 0,$$

if $|y|$ is small enough. Since the slope $\mu : \lambda$ of $\mathfrak{D}^0$ at $(0,0)$ is infinite, $\lambda = 0$ means that $c(0,0) \neq 0$, by (12*). Consequently, (15) holds for

small a, and the solution paths of (3) on $\mathfrak{D}^- + \mathfrak{D}^0$ are defined by (16_1) and (16_2), where $|D| = -D$.

Since (21) is a Jordan arc which divides $\mathfrak{C}_a$ into $\mathfrak{D}^-$ and $\mathfrak{D}^+$, it can, after a rotation by π of the coordinate system (x, y), be assumed that $\mathfrak{D}^-$ (and not $\mathfrak{D}^+$) is to the right of the curve (21); so that the set of points (x, y) contained in $\mathfrak{D}^-$ is characterized by the following pair of conditions:

$$\mathfrak{D}^-: x > x^0(y),\ x^2 + y^2 < a^2 \tag{22}$$

(a is small enough).

6. Let $f(x, y)$ denote either of the functions (16_j) and put

$$f(0, 0) = m. \tag{23}$$

Then any solution of (1) and $y(0) = 0$ lies in the wedge

$$(m - \epsilon)x \leqq y \leqq (m + \epsilon)x \text{ for small } x \geqq 0, \tag{24}$$

where $\epsilon > 0$ can be chosen arbitrarily small. If $f(x, y)$ is defined and continuous on a wedge (24) and satisfies (23), then, according to a standard generalization of the Cauchy-Lipschitz criterion, (1) will possess only one solution $y(x)$ satisfying $y(0) = 0$ if there exists, for small positive x, some (continuous) function $\phi(x)$ satisfying

$$|f(x, u) - f(x, v)| \leqq \phi(x)\,|u - v| \text{ and } \int_{+0} \phi(x)\,dx < \infty, \tag{25}$$

where u and v are arbitrary y-values compatible with (24).

Since (25) is surely true if

$$f(x, u) - f(x, v) = O(x^{-\frac{1}{2}})\,|u - v| \tag{26}$$

holds as $x \to 0$ (with an O which is uniform in u, v), it is sufficient to prove that both functions $f(x, y)$ defined by (16_j) satisfy (26). But the functions $a(x, y)$, $b(x, y)$, $c(x, y)$, $D(x, y)$ are of class C^1. It follows therefore from (15) that it is sufficient to verify the estimate which results from (26) if $f(x, y)$ is replaced by the square root occurring in (16_j); that is, the estimate

$$(-D(x, u))^{\frac{1}{2}} - (-D(x, v))^{\frac{1}{2}} = O(x^{-\frac{1}{2}})\,|u - v|. \tag{27}$$

By the mean-value theorem of differential calculus, the expression on the left of (27) is

$$D(x, v) - D(x, u) \tag{28}$$

times $\frac{1}{2}(-D(x, w))^{-\frac{1}{2}}$, where w is somewhere between u and v. Since (21) and the second of the relations (11) imply that $D_y(0,0)=0$ and $D_x(0,0)\neq 0$ (cf. the arguments at the end of Section **4**), it follows that

$$(29)\qquad D(x,y)=\text{const.}\,x+o(x+|y|)\ \text{ as }\ (x,y)\to(0,0),$$

where const. $=D_x(0,0)\neq 0$. Since (29) remains true if y is replaced by w, a y-value satisfying (24), it follows that

$$(30)\qquad -D(x,w)\geqq \text{Const.}\,x,\ \text{ where Const.}>0.$$

The lower estimate (30), when compared with the representation of the difference (27) as the product mentioned in connection with (28), proves the truth of (27). Hence the proof of (iii) is now complete.

7. Over a domain, say $\mathfrak{C}_a$: $x^2+y^2<a^2$, of the (x,y)-plane, let

$$(31)\qquad \mathfrak{S}\colon\ z=z(x,y)$$

be a surface of class C^3. Then the functions

$$(32)\qquad a=z_{xx},\qquad b=z_{xy},\qquad c=z_{yy}$$

are of class C^1, as is the Gaussian curvature $K=K(x,y)$, the latter being

$$(33)\qquad K=D/(1+z_x{}^2+z_y{}^2)^2,$$

by (4). It will be supposed that

$$(34)\qquad K(0,0)=0\quad\text{and}\quad \operatorname{grad}K(0,0)\neq 0.$$

If the notation is so chosen that the plane tangent to the surface $\mathfrak{S}$ at the point $(0,0)$ is the (x,y)-plane, then, since $z_x(0,0)=0=z_y(0,0)$, it is seen from (33) that (34) is equivalent to (11) in the case (32). The equation (3) becomes that of the asymptotic curves on (31); cf. (32). Accordingly, if $\mathfrak{S}^+$, $\mathfrak{S}^-$, $\mathfrak{S}^0$ denote those portions of the surface $\mathfrak{S}$ which have the respective orthogonal projections $\mathfrak{D}^+$, $\mathfrak{D}^-$, $\mathfrak{D}^0$ in the (x,y)-plane (cf. the beginning of Section **4**), then (33) shows that the situation is as follows:

$\mathfrak{S}$ is divided by a curve $\mathfrak{S}^0$ of class C^1 into two domains, $\mathfrak{S}^+$ and $\mathfrak{S}^-$, the points of which are elliptic $(K>0)$ and hyperbolic $(K<0)$, respectively, while $\mathfrak{S}^0$, which corresponds to the discriminant curve $\mathfrak{D}^0$, is a parabolic curve $(K=0)$. There are no asymptotic curves on $\mathfrak{S}^+$, while those on $\mathfrak{S}^-$ form a net consisting of two distinct families, $\mathcal{F}_1$ and $\mathcal{F}_2$. Finally, each of the latter has the envelope $\mathfrak{S}^0$ provided that the last assumption of (ii), requiring that the parabolic curve $\mathfrak{S}^0$ itself be an asymptotic curve, is

satisfied; whereas there is no envelope and (iii) applies, if (12*) is assumed at the point $(0,0)$ of the projection $\mathfrak{D}^0$ of the parabolic curve $\mathfrak{S}^0$.

It should be noted that these two cases, that of an asymptotic curve $\mathfrak{S}^0$ and that of an $\mathfrak{S}^0$ satisfying (12*), neglect a third possible case. In fact, condition (12), which is the negation of (12*), is necessary but not sufficient in order that $\mathfrak{S}^0$ be an envelope. In other words, it is possible that $\mathfrak{S}^0$ fails to be an asymptotic curve although it starts out at the parabolic point $(0,0)$ in the asymptotic direction of the latter.

The above result should be compared with the following remark of Cohn-Vossen [2], pp. 274-275: The parabolic curve $\mathfrak{S}^0$ is an envelope of the asymptotic curves on $\mathfrak{S}$ if and only if the normal of $\mathfrak{S}$ has a constant direction along $\mathfrak{S}^0$.

In the latter regard, cf. the results of [3], pp. 610-613.

8. The second condition in (11) implies that no point (x,y) of the discriminant curve (5) is a singular point of (3); in other words, that (x,y) is not a common solution of the three equations

$$a(x,y)=0, \qquad b(x,y)=0, \qquad c(x,y)=0. \tag{35}$$

In what follows, (3) will be considered under the assumption that (5) holds on (and only on) an arc $\mathfrak{D}^0$ consisting of singular points (x,y). This is the situation if (3) represents either the differential equation of the lines of curvature on a surface and the set of umbilical points forms an arc $\mathfrak{D}^0$, or the differential equation of asymptotic lines on a surface and the set of flat points forms an arc $\mathfrak{D}^0$.

(iv) *Let the coefficient functions of* (3), *hence also the discriminant* (4), *be of class* C^2 *on a domain, say on* $\mathfrak{C}_a$: $x^2+y^2<a^2$. *Suppose that*

$$|\operatorname{grad} a|+|\operatorname{grad} b|+|\operatorname{grad} c|\neq 0 \tag{36}$$

and that there is through $(0,0)$ *an arc* $\mathfrak{D}^0$ *along which* (35) *holds (so that* $\mathfrak{D}^0$ *is an arc of class* C^2). *Suppose further that*

$$D(x,y)\lesseqgtr 0 \text{ according as } (x,y) \text{ is not or is on } \mathfrak{D}^0, \tag{37}$$

and that

$$D(x,y)\neq o(r^2), \quad \text{as} \quad r^2=x^2+y^2\to 0. \tag{38}$$

Then the solution paths of (3) *form two disjoint families* $\mathfrak{F}_1$, $\mathfrak{F}_2$ *each of which covers* $\mathfrak{C}_a$ *in a schlicht fashion (in fact, every point* (x,y) *of* $\mathfrak{C}_a$ *issues exactly two solution paths of* (3), *and these paths are not tangent to each other).*

Thus there is no envelope even though the discriminant curve is a "line of singular points."

In the proof of (iv), it will be shown that (3) is in the main equivalent to a non-singular differential equation (49), one for which the corresponding discriminant is not 0. In the assertion of (iv), the curve $\mathfrak{D}^0$ is considered a solution of (3) if and only if it is a solution of (49).

Proof of (iv). After a suitable rotation of the (x, y)-plane, it can be supposed that $\mathfrak{D}^0$ is tangent to the x-axis at $(0, 0)$. Since (35) holds along $\mathfrak{D}^0$, it follows from (36) that there exist three constants c_1, c_2, c_3, not all 0, such that

$$(39) \qquad a = c_1 y + f_1, \qquad b = c_2 y + f_2, \qquad c = c_3 y + f_3,$$

where $f_k = f_k(x, y)$ is a function of class C^2 satisfying

$$(40) \qquad f(x, y) = o(r) \quad \text{as} \quad r \to 0.$$

Conditions (37) and (38) imply that

$$(41) \qquad c_1 c_3 - c_2^2 < 0.$$

For small $|x|$, the discriminant curve $\mathfrak{D}^0$ is of the form (14). In order to reduce this to the simpler case

$$(42) \qquad \mathfrak{D}^0 \colon\; y \equiv 0,$$

replace (x, y) by $(x, y - y^0(x))$ and denote $y - y^0(x)$ simply by y. Then (3) goes over into an equation of the type

$$(43) \qquad y(A dx^2 + 2B dx dy + C dy^2) = 0,$$

if

$$(44) \qquad yA = a + 2by^{0\prime} + cy^{0\prime 2}, \qquad yB = b + cy^{0\prime}, \qquad yC = c,$$

where the argument of a, b, c is $(x, y + y^0(x))$ and that of $y^{0\prime}$ is x. Thus the functions (44) (of the new (x, y)), as well as their partial derivatives with respect to y, are of class C^1 (since y^0 is of class C^2); moreover, the functions (44) (of the new (x, y)) are of the form

$$(45) \qquad yA = c_1 y + F_1, \qquad yB = c_2 y + F_2, \qquad yC = c_3 y + F_3,$$

where $F_k(x, y)$ and $\partial F_k(x, y)/\partial y$ are of class C^1, and

$$(46) \qquad F_k(x, y) = o(r) \quad \text{and} \quad F_k(x, 0) \equiv 0.$$

It follows that A, B, C are of class C^1 and satisfy

$$(47) \qquad A(0, 0) = c_1, \qquad B(0, 0) = c_2, \qquad C(0, 0) = c_3.$$

Hence, by (41),

$$(48) \qquad AC - B^2 < 0$$

in a sufficiently small vicinity of $(x, y) = (0, 0)$. Consequently, (1) splits into the two equations, $y = 0$ and

$$(49) \qquad Adx^2 + 2Bdxdy + Cdy^2 = 0.$$

Hence the assertion of (iv) follows from (48) and the fact that A, B, C are of class C^1.

Remark. If the C^2-assumption of (iv) is weakened to the condition that a, b, c are of class C^1, then $\mathfrak{D}^0$ *can* be an envelope. This is shown by the example

$$|\,y\,|^{3/2}dx^2 - (y + |\,y\,|^{3/2})dxdy + ydy^2 = 0.$$

In fact, this case of (3) factors into

$$y = 0, \qquad dy/dx = |\,y\,|^{\frac{1}{2}} \operatorname{sgn} y, \qquad dy/dx = 1;$$

hence $\mathfrak{D}^0$: $y \equiv 0$ is an envelope, although conditions (36), (37), (38) are satisfied.

9. It turns out that even if (38) is not assumed, so that

$$(50) \qquad D(x, y) = o(r^2),$$

the discriminant curve $\mathfrak{D}^0$ is still not an envelope. But the assertion of (iv) need not of course hold in the case (50).

(v) *Let all conditions of* (iv) *except* (38) *be satisfied and, in addition to* (50), *suppose that a, b, c are of class C^3. Then $\mathfrak{D}^0$ is not an envelope of solutions of* (3).

Proof of (v). The considerations occurring in the proof of (iv) are valid except that, (38) being negated, (41) and (45) do not hold. Thus

$$(51) \qquad c_1c_3 - c_2{}^2 = 0$$

and

$$(52) \qquad AC - B^2 = 0$$

hold for $y = 0$. On the other hand, since condition (36) has been retained,

$$(53) \qquad |c_1| + |c_2| + |c_3| \neq 0.$$

Suppose, if possible, that $\mathfrak{D}^0$ is an envelope of solutions of (3). Then (42) is an envelope of solutions of (49). It will be shown that this leads to a contradiction.

Since $y = 0$ is an envelope of solutions of (49), the function $y \equiv 0$ is a solution of (49). Hence $A(x, 0) \equiv 0$, that is, $c_1 = 0$. But then (51) and (53) imply the truth of the last two of the three relations

$$(54) \qquad c_1 = 0, \qquad c_2 = 0, \qquad c_3 \neq 0.$$

Thus, for (x, y) near $(0, 0)$, (49) can be written as

$$(55) \qquad dy/dx = \{-B \pm (B^2 - AC)^{\frac{1}{2}}\}/C, \qquad (C \neq 0),$$

where

$$(56) \qquad B(x, 0) \equiv 0.$$

The assumption (37) implies that $AC - B^2 \lesseqgtr 0$ according as $|y| \gtreqless 0$. Hence, $\partial(AC - B^2)/\partial y = 0$ at $(x, y) = (x, 0)$, and so, since a, b, c are of class C^3,

$$A(x, y)C(x, y) - B^2(x, y) = O(y^2),$$

as $y \to 0$, uniformly in x for small $|x|$. Consequently, (54) and (56) imply that

$$|dy/dx| = O(|y|)$$

holds for any solution path of (55). Accordingly $y \equiv 0$ is the only solution of (55) satisfying $y(x_0) = 0$ (if $|x_0|$ is small enough). Hence $y \equiv 0$ is not an envelope of solutions of (55). This proves (v).

10. Over a domain, say $\mathfrak{C}_a$: $x^2 + y^2 < a^2$, of the (x, y)-plane, let (31) be a surface of class C^4. Then, if $p = p(x, y), \cdots, t = t(x, y)$ denote the partial derivatives $z_x, \cdots, z_{yy}$, the functions

$$(57) \qquad a = pqr - (1 + p^2)s, \qquad 2b = (1 + q^2)r - (1 + p^2)t,$$
$$c = (1 + q^2)s - pqt$$

are of class C^2. For the choice (57) of the coefficient functions, (3) represents the differential equations of the lines of curvature on $\mathfrak{S}$.

The Gaussian curvature $K = K(x, y)$ and mean curvature $H = H(x, y)$ of $\mathfrak{S}$ satisfy the inequality $K - H^2 \leqq 0$. An umbilical point is characterized by the sign of equality, $K - H^2 = 0$. At such a point (x, y), the functions (57) satisfy (35).

If $(x, y) = (0, 0)$ is an umbilical point, the axes x, y, z can be chosen so that

$$z(x, y) = \tfrac{1}{2}K(0, 0)(x^2 + y^2) + \psi(x, y)/6 + o(r^4) \text{ as } r \to 0,$$

where $\psi(x, y)$ is a cubic form,

$$\psi(x, y) = a_{30}x^3 + 3a_{21}x^2y + 3a_{12}xy^2 + a_{03}y^3.$$

It is readily verified from (57) that

$$-\operatorname{grad} a = (a_{21}, a_{12}) = \operatorname{grad} c \text{ and } 2 \operatorname{grad} b = (a_{30} - a_{12}, a_{03} - a_{21})$$

at $(x, y) = (0, 0)$. Hence (36) holds at $(0, 0)$ if and only if

$$| a_{21} | + | a_{12} | + | a_{30} - a_{12} | + | a_{03} - a_{21} | \neq 0,$$

which means that $\psi(x, y) \not\equiv 0$. Since $\operatorname{grad} a = -\operatorname{grad} c$ at $(x, y) = (0, 0)$, it follows that (38) is satisfied.

Accordingly, (iv) implies that if $\psi(x, y) \not\equiv 0$ and if the set of umbilical points $(H^2 - K = 0)$ consists of an arc $\mathfrak{D}^0$ through $(0, 0)$ (hence $\mathfrak{D}^0$ is an arc of class C^2), then the set of lines of curvature form two disjoint families $\mathfrak{F}_1$, $\mathfrak{F}_2$, each of which affords a *schlicht* covering of a vicinity of $(0, 0)$.

In this assertion, $\mathfrak{D}^0$ is counted at most once as a line of curvature; cf. the remark following (iv).

Appendix.

In a neighborhood, $\mathcal{V}$, of $(x, y) = (0, 0)$, let $\gamma = \gamma(x, y)$, where $\gamma(0, 0) = 0$, be a continuous function having the following properties: If the value of the constant in the equation $\gamma(x, y) = \text{const.}$ is within a certain interval, then the equation represents a Jordan arc *possessing a continuous tangent*, and varying with the value const. continuously and in such a way that all Jordan arcs together cover $\mathcal{V}$ in a *schlicht* continuous fashion. Then $\gamma(x, y) = \text{const.}$ will be called a C^0-*sheaf* (on $\mathcal{V}$). If, in addition, $\gamma(x, y)$ is a function of class C^1 and of non-vanishing gradient $(\gamma_x{}^2 + \gamma_y{}^2 \neq 0)$, then the C^0-sheaf will be called a C^1-*sheaf* (on some, sufficiently small, neighborhood $\mathcal{V}$ of the origin).

If $\gamma = \alpha = \alpha(x, y)$ is of class C^1 and of non-vanishing gradient, then a classical rule would state that the orthogonal trajectories of the C^1-sheaf $\alpha(x, y) = \text{const.}$ are given by the sheaf $\gamma(x, y) = \text{Const.}$ belonging to $\gamma = \beta$, where $\beta = \beta(x, y)$ is any non-constant solution (of class C^1?) of the partial differential equation

$$\alpha_x \beta_x + \alpha_y \beta_y = 0. \tag{1}$$

The following pair of assertions shows, however, that this formulation of the statement is both misleading and false.

(*) *If $\alpha(x, y)$ is a function of class C^1 and of non-vanishing gradient in a neighborhood of the point $(0, 0)$, then the C^1-sheaf, defined by $\alpha(x, y) = \text{const.}$ (for sufficiently small $x^2 + y^2$),*

(I) *need not have any C^0-sheaf, say $\beta(x, y) = \text{Const.}$, of orthogonal trajectories (on any neighborhood $x^2 + y^2 < \epsilon^2$);*

(II) *can have a C^0-sheaf of orthogonal trajectories on some, without having a C^1-sheaf of orthogonal trajectories on any, neighborhood of $(x, y) = (0, 0)$.*

In view of the criterion (1), both of these assertions, (I) and (II), can be interpreted as statements concerning possibilities which can occur for a partial differential equation

$$a(x, y)\beta_x + b(x, y)\beta_y = 0, \tag{2}$$

for which a non-constant solution $\beta(x, y)$ is sought when $a(x, y)$, $b(x, y)$ are the components of the gradient of a given function, $\alpha(x, y)$, which is of class C^1 and of non-vanishing gradient but otherwise arbitrary. Such possibilities cannot of course occur if grad α satisfies a Lipschitz condition. In the examples to be given below, grad α will satisfy a Hölder condition, of an index $\lambda (< 1)$ which can be chosen arbitrarily close to Lipschitz's limiting case $\lambda = 1$.

Proof of (*). Both (I) and (II) will be proved by functions $\alpha(x, y)$ of the form

$$\alpha(x, y) = x + f(y), \text{ hence } \alpha_x = 1, \alpha_y = f', \tag{3}$$

where $f' = df/dy$, and by considering the differential equation

$$dy/dx = -f'(y), \tag{4}$$

which, in view of (3), defines the characteristics of (1). In terms of a constant λ satisfying $0 < \lambda < 1$, the function $f(y)$ will be defined to be

(5_{I}) $\quad f(y) = |\,y\,|^{1+\lambda};$ $\qquad\qquad$ (5_{II}) $\quad f(y) = |\,y\,|^{1+\lambda}\operatorname{sgn} y$

in the proof of the respective assertions (I), (II). In both cases, (3) shows that $\alpha(x,y)$ is of class C^1 and that $\alpha(0,0) = 0$, $\operatorname{grad}\alpha(0,0) \neq 0$.

In both cases, it is clear from (3) that if $\beta(x,y) =$ Const. is a C^1-sheaf of orthogonal trajectories of the C^1-sheaf $\alpha(x,y) =$ const., then $\beta(x, y(x))$ must be independent of x along every solution $y = y(x)$ of the differential equation (4). But the latter is

(6_{I}) $\quad dy/dx = -(1+\lambda)\,|\,y\,|^{\lambda}\operatorname{sgn} y;$ $\qquad$ (6_{II}) $\quad dy/dx = -(1+\lambda)\,|\,y\,|^{\lambda}$

in the respective cases (5_{I}), (5_{II}). Since the x-axis represents a solution of both (6_{I}) and (6_{II}), and since it is clear that every other solution curve must meet the x-axis, it follows first, that $\beta(x,0)$ is independent of x, and then, that $\beta(x,y)$ is independent of both x and y.

This proves that in neither case can there exist a C^1-sheaf of orthogonal trajectories. In the second case, it is readily verified from (5_{II}) that if $\beta(x,y)$ is defined by

$$(7_{\mathrm{II}}) \qquad \beta(x,y) = (1+\lambda)(1-\lambda)x + |\,y\,|^{1-\lambda}\operatorname{sgn} y,$$

then $\beta(x,y) =$ Const. is a C^0-sheaf of orthogonal trajectories. This proves assertion (II) of (*).

In order to prove assertion (I) of (*), suppose, if possible, that there exists a C^0-sheaf, $\beta(x,y) =$ Const., in a neighborhood of the point $(x,y) = (0,0)$ in the case (6_{I}). Within the first quadrant $(x > 0, y > 0)$, let (x_0, y_0) be any point situated on the curve

$$(7_{\mathrm{II}}) \qquad (1+\lambda)(1-\lambda)x - |\,y\,|^{1-\lambda} = 0.$$

There is through (x_0, y_0) a unique path, say $\Gamma^+ = \Gamma^+(x_0, y_0)$, which is an orthogonal trajectory of class C^1, this Γ^+ being represented by the *positive* solution $y = y(x)$ of the equation (7_{II}) or by the half-line $y(x) \equiv 0$ according as $0 < x < \infty$ or $-\infty < x \leqq 0$. Hence $\beta = \beta(x,y)$ is constant along Γ^+. Similarly, $\beta = \beta(x,y)$ is constant along Γ^- if, according as $-\infty < x \leqq 0$ or $0 < x < \infty$, the path $\Gamma^- = \Gamma^-(x_0, -y_0)$ is defined to be the half-line $y(x) \equiv 0$ or the negative solution $y = y(x)$ of the equation (7_{II}) which passes through the point $(x_0, -y_0)$ of the fourth quadrant $(x_0 > 0, -y_0 < 0)$. In other words, Γ^- is the unique path of class C^1 which passes through $(x_0, -y_0)$ and is an orthogonal trajectory. Since (7_{II}) remains unchanged if y is replaced

by $-y$, the curve Γ^- results from the curve Γ^+ by reflection on the x-axis. But Γ^+ and Γ^- have the negative half of the x-axis in common. Since the function $\beta(x, y)$ is constant on Γ^+ as well as on Γ^-, it follows that there exists a number, say c_0, satisfying $\beta(x, y) = c_0$ for every point (x, y) of $\Gamma^+ + \Gamma^-$. This completes the proof, since $\beta(x, y) = c_0$ fails to define an arc of class C^1 through $(0, 0)$ (simply because Γ^+ and Γ^-, which are distinct for $x > 0$, $\pm y > 0$, coalesce for $x \leqq 0, y = 0$).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. L. Cauchy, *Leçons de calcul intégral*, vol. 2, Paris, 1844.

[2] S. Cohn-Vossen, "Die parabolische Kurve," *Mathematische Annalen*, vol. 99 (1928), pp. 273-308.

[3] P. Hartman and A. Wintner, "On geodesic torsions and parabolic and asymptotic curves," *American Journal of Mathematics*, vol. 72 (1952), pp. 607-625.

[4] W. F. Osgood, "Beweis der Existenz einer Lösung der Differentialgleichung $y' = f(x, y)$," *Monatshefte für Mathematik und Physik*, vol. 9 (1898), pp. 331-345.

[5] O. Perron, "Über Differentialgleichungen erster Ordnung, die nicht nach der Ableitung aufgelöst sind," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 22 (1913), pp. 356-638.

[6] S. Wallach, "The differential equation $y' = f(y)$," *American Journal of Mathematics*, vol. 70 (1948), pp. 345-350.

ON THE EQUATION $a^x - b^y = 1$.*

By J. W. S. Cassels.

1. The solution of the title equation in integers x, y for given integers $a > 1, b > 1$ has been dicussed by W. J. LeVeque [2]. The second paragraph of this note shows that y is odd if $x > 1$, and that any prime divisor of y which is less than x divides a, and *vice versa.* The third paragraph proves simply a stronger form of LeVeque's theorem, that there is at most one solution, which can be specified completely. The third paragraph uses the results of the second only to secure a slight refinement of the enunciation.

It is conjectured that there are only a finite number of nontrivial solutions a, b, x, y of the equation.

2. We first have the trivial

Theorem I. $(x, y) = 1$.

Proof. Suppose that $x = px_1$, $y = py_1$ for $p > 1$. Then $a_1{}^p - b_1{}^p = 1$ with $a_1 = a^{x_1}$, $b_1 = b^{y_1}$; which is clearly impossible.

Theorem II. *If* $x > 1$, *then* $2 \nmid y$.

Proof. Otherwise, by the preceding argument we should have a solution a, b, x, y of $a^x - b^2 = 1$ with odd prime x. But then $1 + ib = \epsilon(p + iq)^x$, $p^2 + q^2 = a$ for some unit ϵ and by replacing $p + iq$ by $\eta(p + iq)$ with a suitable unit η we may suppose that $\epsilon = 1$. Equating coefficients we now have

$$1 = p(p^{x-1} - \tfrac{1}{2}x(x-1)p^{x-3}q^2 + \cdots \pm xq^{x-1}),$$

and so $p = \pm 1$. By considering conguences modulo x we have p $= 1$, and so

$$(x-1)/2 - (x-1)(x-2)(x-3)q^2/4! + \cdots \pm q^{x-1} = 0. \tag{1}$$

Since $2 \mid (1 \pm i)^x$ we have $| q | > 1$. Let r be a prime divisor of q. We shall show that all the terms on the left of (1) except the first are

* Received September 8, 1952.

divisible by a higher power of r than that dividing $(x-1)/2$; which contradicts (1). It is enough to show that for $k \geqq 2$ the fraction

$$2(x-2)\cdots(x-2k+1)q^{2k-3}/(2k)!$$
$$=(x-2)\cdots(x-2k+1)/(2k-2)!\cdot q^{2k-3}/k(2k-1)$$

does not have r in its denominator when reduced. The first factor is an integer. For $k \geqq 4$ we have $r^{2k-3} \mid q^{2k-3}$, but $r^{2k-3} \geqq 2^{2k-3} > k(2k-1)$, so then the statement is certainly true. For $k = 2, 3$ we have $r \mid q^{2k-3}$, but $r^2 \nmid k(2k-1)$ since $k(2k-1) = 6, 15$ respectively is squarefree, and again the statement is true. Hence the assumption that $a^x - b^2 = 1$ is soluble leads to a contradiction.

We require two trivial lemmas.

LEMMA 1. *Let p be an odd prime and $c > 1$ an integer. Then $f = (c^p - 1)/(c-1)$ is prime to p or divisible by p but not by p^2 according as $c \not\equiv 1\ (p)$ or $c \equiv 1\ (p)$. The number $f, f/p$ respectively is odd, greater than 1 and prime to $c-1$.*

Further, $g = (c^p + 1)/(c+1)$ is prime to p or divisible by p but not by p^2 according as $c \not\equiv -1\ (p)$ or $c \equiv -1\ (p)$. The number $g, g/p$ respectively is odd and prime to $c+1$; it is greater than 1 except when

$$(2) \qquad c = 2, \qquad p = 3.$$

Proof. If q is a prime divisor of $c-1$, then $f = 1 + c + \cdots c^{p-1} \equiv p\ (q)$ and so $q \mid f$ implies $q = p$. If $c = 1 + rp$, then

$$f \equiv 1 + (1+rp) + (1+2rp) + \cdots + (1+(p-1)rp) \equiv p\,(p^2)$$

and so the greatest common divisor of $c-1$, f is 1 or p. In particular f is odd if c is odd. If c is even, then $f \equiv 1/1\ (2)$ is again odd. Finally, it is obvious that $f > p$.

As before, the greatest common divisor of g and $c+1$ is 1 or p, and g is odd. Also

$$\frac{g}{p} = \frac{c^p+1}{p(c+1)} \geqq \frac{2^p+1}{p.3} \geqq \frac{2^3+1}{3.3} = 1,$$

with equality in both places only if $c = 2$, $p = 3$.

LEMMA 2. *Let $c > 1$. If c is even, then $c+1$, $c-1$ are coprime. If c is odd, then one of $c+1$, $c-1$ say $c \pm 1$, is not divisible by 4; and then $\frac{1}{2}(c \pm 1)$ is prime to $c \mp 1$.*

Proof. Clear

We can now prove

THEOREM III. (i) *If p is prime and $p \mid x$, $p \nmid b$, then $p > y$.*

(ii) *If p is prime and $p \mid y$, $p \nmid a$, then $p > x$.*

Proof. We first prove (i) and put $x = px_1$. By Lemmas 1, 2 the numbers $(a^x - 1)/(a^{x_1} - 1)$ and $a^{x_1} - 1$ are coprime, and so $a^{x_1} - 1 = c^y$ for some $c \mid b$. Hence $b^y = (c^y + 1)^p - 1$ and so $b > c^p$, i. e. $b \geqq c^p + 1$. Then

$$(c^p + 1)^y \leqq b^y = (c^y + 1)^p - 1 < (c^y + 1)^p,$$

and so $p > y$ ([1] Theorem 19).

For (ii) we first note that $p > 2$ by Theorem II. Put $y = py_1$ and so, as before, $b^{y_1} + 1 = d^x > 1$ for some $d \mid a$. Hence $a^x = (d^x - 1)^p + 1$ and so $a \leqq d^p - 1$. Thus

$$(d^p - 1)^x \geqq a^y = (d^x - 1)^p + 1 > (d^x - 1)^p,$$

and so $p > x$.

We call a solution nontrivial if $x > 1$, $y > 1$ and deduce

COROLLARY 1. *For a non-trivial solution it is impossible that*

$$(x, b) = (y, a) = 1.$$

For a later purpose we require

COROLLARY 2. *There are no nontrivial solutions of $2^x - b^y = 1$.*

Proof. If $y > 1$, $b > 1$ then $x > 1$ and so y is odd by Theorem II. Hence each prime factor of y is greater than x and in particular $b^y > 2^y > 2^x$, a contradiction.

3. The following theorem enables all solutions of the title equation to be found for given a, b.

THEOREM IV. *Let*

$$a^x - b^y = 1, \tag{3}$$

where $x, y, a > 1$, $b > 1$ are positive integers and the equation is not

$$3^2 - 2^3 = 1. \tag{4}$$

Suppose that ξ, η are the least positive solutions of

$$a^{\xi} \equiv 1 \ (B), \qquad b^{\eta} \equiv -1 \ (A),$$

where A, B are the products of the odd primes dividing a, b respectively.

Then $x = \xi$, $y = \eta$; except that $x = 2$, $y = 1$ may occur if $\xi = \eta = 1$ and $a + 1$ is a power of 2.

Proof. We first prove $y = \eta$. Clearly $\eta \mid y$. Suppose y/η is even. Then $b^y \equiv (-1)^{y/\eta} \equiv 1 \not\equiv -1 \ (A)$ unless $A = 1$, i. e. unless a is a power of 2. But then $a^x = b^y + 1 \equiv 2 \ (4)$ and so $a^x = 2$, $b = 1$; which is excluded. Hence y/η is odd. Suppose that y/η is divisible by an odd prime p, say $y = py_1, \eta \mid y_1$. Then by the second part of Lemma 1 there is an odd prime q dividing $(b^y + 1)/(b^{y_1} + 1)$ (and so a) but not dividing $b^{y_1} + 1$; except in the case (2) which corresponds to (4). Hence $b^{y_1} + 1 \not\equiv 0 \ (q)$ and a fortiori $b^{y_1} \not\equiv -1 \ (A)$. The contradiction proves $y = \eta$.

We now prove the statements about x. Clearly $\xi \mid x$. The proof that x/ξ is a power of 2 runs exactly as before using now the first part of Lemma 1. If $2\xi \mid x$, say $x = 2x_1$, $\xi \mid x_1$, then a similar argument using Lemma 2 leads to an absurdity unless $a^{x_1} + 1$ contains no odd prime factors, i. e. $a^{x_1} + 1 = 2^m$ for some $m > 0$. If now $x_1 \neq \xi$, then $2 \mid x_1$ and so $2^m = a^{x_1} + 1 \equiv 2 \ (4)$ i. e. $m = a = 1$, which is excluded.

Hence $x = \xi$ or $x = 2\xi$, the latter only if

$$a^{\xi} + 1 = 2^m. \tag{5}$$

But (5) implies $\xi = 1$ by Theorem III, Corollary 2. Now $a + 1 = 2^m$, $a^2 - 1 = b^y$ and hence $a - 1 = 2c^y$ for some odd c, where $y \mid (m + 1)$. Finally, $2 = 2^m - 2c^y$ and hence $1 = 2^{m-1} - c^y$. By Theorem III, Corollary 2 this implies $c = 1$ or $y = \eta = 1$. The case $c = 1$ gives $a = 3$ and so the exception (4) of the theorem; and the case $y = \eta = 1$ gives the exception at the end of the enunciation.

TRINITY COLLEGE, CAMBRIDGE, ENGLAND.

REFERENCES.

[1] G. H. Hardy, J. E. Littlewood and G. Polya, *Inequalities*, Cambridge, 1934.

[2] W. J. LeVeque, "On the equation $a^x - b^y = 1$," *American Journal of Mathematics*, vol. 74 (1952), pp. 325-331.

SOME CONGRUENCES FOR THE BERNOULLI NUMBERS.*

By L. CARLITZ.

1. Introduction. The familiar Kummer congruences [2] for the Bernoulli numbers may be written in the form

$$\tau^m(\tau^{p-1}-1)^r \equiv 0 \pmod{p^r}, \tag{1.1}$$

where after expansion of the left member, τ^m is to be replaced by $\tau_m = B_m/m$, and $(B+1)^m = B^m$ for $m \neq 1$; p is a prime > 2; $m \geqq r+1$ and $p-1 \nmid m$. Somewhat more generally, if $(p-1)p^{e-1} \mid b$, then, by [1], p. 842,

$$\tau^m(\tau^b-1)^r \equiv 0 \pmod{p^{re}}, \tag{1.2}$$

provided $p > 2$, $p-1 \nmid m$, $m > re$. In attempting to remove the condition $p-1 \nmid m$, Vandiver [5] proved the following result:

$$B^{a(p-1)}(B^{p-1}-1)^r \equiv 0 \pmod{p^{r-1}}, \tag{1.3}$$

where $a > 0$, $r > 0, a+r < p-1$.

Vandiver's conguence differs from (1.1) in several respects; the most striking difference is the restriction on r. In the present paper we shall extend (1.3) in several directions. We prove first that

$$\sigma^a(\sigma-1)^r \equiv 0 \pmod{p^r}, \tag{1.4}$$

where $\sigma_m = (B_{m(p-1)} + p^{-1} - 1)/m$, $a \geqq 1$, $r \geqq 1$, $a+r \leqq p-1$; next if $m = (hp+k)(p-1)$, $h \geqq 0$, $k \geqq 1$, $r \geqq 1$, $k+r \leqq p-1$, then

$$B^m(B^{p-1}-1)^r \equiv 0 \pmod{p^{r-1}}. \tag{1.5}$$

These results can however be extended considerably. We first show that if $p^r \mid m$, then σ_m is integral $(\bmod p)$, in other words the numerator of $B_{m(p-1)} + p^{-1} - 1$ is divisible by p^r; moreover we determine the residue of $\sigma_m \pmod{p}$. In the next place we show that

$$B^c(B^b-1)^r \equiv 0 \pmod{p^{re-h}}, \tag{1.6}$$

where $(p-1)p^{e-1} \mid b$, $c = (p-1)u > re$, and $h = e$ for $r < p$ (except perhaps when $r = p-1$, $e = 1$ and $h = 2$), for $r \geqq p$, h is the least integer $\geqq (re+1)/p$.

---

* Received November 20, 1951.

Finally we quote

$$\sum_{s=0}^{r} (-1)^{r-s} C_s{}^r (s+1)^{-1}(B_m + p^{-1} - 1) \equiv 0 \pmod{p^{b-h},\ (r+1)^{-1}p^{e(r+1)-h}} \tag{1.7}$$

where $(p-1)p^{e-1} \mid b$, $m = (s+1)b$, and $h = e$ for $r < p$, while for $r > p$, h is the least integer $\geqq re/p$; in the modulus it is to be understood that we are interested only in the power of p dividing each term.

For a result more general than (1.7) see Theorem 6.

2. Some preliminaries. We shall require an extension of a formula used by Nielsen ([4], p. 266). Put

$$A_{k,r} = \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r C_k{}^{sb+t} U_{sb+m}, \tag{2.1}$$

where $r \geqq 0$, $k \geqq 0$. Then it is readily verified by direct substitution that

$$(t + rb - k)_{k,r} + rbA_{k,r-1} = (k+1)A_{k+1,r}. \tag{2.2}$$

Now in the first place, if we take $U_m = \tau_m$, $t = m$, it follows from (1.2) and (2.2) that

$$B^m(B^b - 1)^r \equiv 0 \pmod{p^{(r-1)e}} \tag{2.3}$$

provided $p - 1 \nmid m$, $m > re$. In the second place, if we take $U_m = B_m$, then (2.2) and (2.3) imply (since the left member of (2.3) is $A_{0,r}$)

$$\sum_{s=0}^{r} (-1)^{r-s} C_s{}^r C_k{}^{sb+t} B_{bs+m} \equiv 0 \pmod{p^{(r-k-1)e}}, \tag{2.4}$$

provided $p - 1 \nmid m$, $m > re$, $k < p$; the integer t is arbitrary.

3. Proof of (1.4) and (1.5). Put

$$a^{p-1} - 1 = pq_a \qquad (p \nmid a), \tag{3.1}$$

so that q_a is an integer. Then raising both members of (3.1) to the $(r+1)$-th power and summing over a from 1 to $p-1$, we get

$$\sum_{s=1}^{r+1} (-1)^{r+1-s} C_s{}^{r+1} (S_{s(p-1)} + 1 - p) = p^{r+1} \sum_{a=1}^{p-1} q_a{}^{r+1} \tag{3.2}$$

where $S_m = S_m(p) = \sum_{a=1}^{p-1} a^m$. In place of (3.2) it will be more convenient to write

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r(s+1)^{-1}(S_m+1-p)=(r+1)^{-1}p^{r+1}\sum_{a=1}^{p-1}q_a^{r+1} \quad (m=(s+1)(p-1)), \tag{3.3}$$

which is an immediate consequence of (3.2). We next recall that

$$S_m(p)=(m+1)^{-1}\sum_{t=1}^{m+1}C_t^{m+1}B_{m+1-t}p^t. \tag{3.4}$$

Now for $r\leqq p-1$, $t\geqq p$, the t-th term in the right member of (3.4) is divisible by at least p^{p-1}. Thus (3.4) implies

$$S_m\equiv\sum_{t=0}^{p-2}(t+1)^{-1}C_t^mB_{m-t}p^{t+1} \pmod{p^{p-1}}. \tag{3.5}$$

Substituting in the left member of (3.3) we get

$$\begin{aligned}&\sum_{s=0}^{r}(-1)^{r-s}C_s^r(s+1)^{-1}(pB_m+1-p)\\ &\quad+(p-1)\sum_{t=1}^{p-2}t^{-1}(t+1)^{-1}p^{t+1}\sum_{s=0}^{r}(-1)^{r-s}C_s^r\,C_{t-1}^{m-1}B_{m-t}\\ &\qquad\equiv(r+1)^{-1}p^{r+1}\sum_{a=1}^{p-1}q_a^{r+1} \pmod{p^{p-1}},\end{aligned} \tag{3.6}$$

where again $m=(s+1)(p-1$. Suppose now that $r<p-1$ and apply (2.4) with $b=p-1$. Then

$$p^{t+1}\sum_{s=0}^{r}(-1)^{r-s}C_s^rC_{t-1}^{m-1}B_{m-t}\equiv 0$$

mod p^{r+1} or mod p^{t+1} according as $r\geqq t$ or $t>r$; so that the left member is always divisible by p^{r+1} and thus the double sum in the left member of (3.6) is divisible by p^{r+1}. Hence it follows from (3.6) that

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r\sigma_{s+1}\equiv 0 \pmod{p^r} \qquad (r<p-1), \tag{3.7}$$

where

$$\sigma_m=m^{-1}(B_{m(p-1)}+p^{-1}-1). \tag{3.8}$$

In the standard notation of finite differences, we have

$$\Delta^r\sigma_{a+1}=\sum_{s=0}^{r}(-1)^{r-s}C_s^r\sigma_{s+a+1}=\sum_{i=0}^{a}C_i^a\Delta^{r+i}\sigma_1;$$

since the left member of (3.7) is evidently $\Delta^r\sigma_1$, it follows that $\Delta^r\sigma_{a+1}\equiv 0 \pmod{p^r}$ provided $a+r<p-1$. Changing the notation slightly, we now state

THEOREM 1. *If* $a \geqq 1$, $r \geqq 1$, $a + r \leqq p - 1$, *then*

$$\sigma^a(\sigma - 1)^r \equiv 0 \pmod{p^r}, \tag{3.9}$$

where σ_m *is defined by* (3.8).

In the next place, if we raise both members of (3.1) to the r-th power and multiply by $a^{t(p-1)}$, we get

$$\sum_{s=0}^{r} (-1)^{r-s} C_s{}^r S_{(s+t)(p-1)} = p^r \sum_{a=1}^{p-1} a^{t(p-1)} q_a{}^r.$$

We now suppose $t = hp + k$, $k \geqq 1$, $k + r \leqq p - 1$. We again use (3.4) and break the sum into two parts. As before (3.5) holds and we get in place of (3.6):

$$p \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r B_m + \sum_{t=1}^{p-2} (t+1)^{-1} p^{t+1} \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r C_t{}^m B_{m-t} \equiv p^r \sum_{a=1}^{p-1} a^{t(p-1)} q_a{}^r \pmod{p^{p-1}}, \tag{3.10}$$

where $m = (s + t)(p - 1)$. As in the proof of (3.7), the double sum in (3.10) is divisible by p^r, and it follows that

$$p \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r B_m \equiv 0 \pmod{p^r}.$$

We have proved

THEOREM 2. *If*

$$m = (hp + k)(p - 1),\ h \geqq 0,\ k \geqq 1,\ r \geqq 1,\ k + r \leqq p - 1,$$

then

$$B^m(B^{p-1} - 1)^r \equiv 0 \pmod{p^{r-1}}. \tag{3.11}$$

Some of the restrictions in this theorem can be removed; see Theorem 4 below.

4. The numerator of $B_m + p^{-1} - 1$. Let $m = r(p-1)p^k$, so that $a^m \equiv 1 + rp^{k+1} q_a \pmod{p^{k+2}}$, where $a^{p-1} = 1 + pq_a$, $p \nmid a$. Then

$$S_m = \sum_{a=1}^{p-1} a^m \equiv p - 1 + rp^{k+1} w_p \pmod{p^{k+2}}, \tag{4.1}$$

since $\sum_{a=1}^{p-1} q_a \equiv w_p \equiv p^{-1}((p-1)! + 1) \pmod{p}$; cf. [4], p. 356 or p. 361. On the other hand, it follows from (3.4) that

$$S_m \equiv pB_m \pmod{p^{k+2}}, \tag{4.2}$$

provided $p > 3$. Combining (4. 2) with (4. 1) we get $B_m + p^{-1} - 1 \equiv rp^k w_p$ $(\mathrm{mod}\, p^{k+1})$ and therefore

$$(4.3) \qquad p^{-k}(B_m + p^{-1} - 1) \equiv rw_p \pmod{p} \qquad (p > 3).$$

THEOREM 3. *For $m = rp^k(p-1)$, $p \geqq 3$, the numerator of $B_m + p^{-1} - 1$ is divisible by p^k, and the quotient satisfies* (4. 3), *where w_p denotes Wilson's quotient.*

For $k = 0$, see [3], p. 354. By the method used below we can prove the more precise result

$$(4.4) \quad p^{-k}(B_m + p^{-1} - 1) \equiv p^{-h-k}(S_m(p^h) + p^{h-1} - p^h)$$
$$\equiv rp^{1-h} \sum_{a=1}^{p^h}{}' q_k(a) \pmod{p^{2h}},$$

where $p \nmid a$ in the summation $\sum'$ and $h = [\frac{1}{3}(k+2)]$, $S_m(p^h) = \sum_{a=1}^{p^h-1} a^m$, $q_k(a) = p^{-(k+1)}(a^{(p-1)p^k} - 1)$. We shall omit the proof of (4. 4).

5. Proof of (1. 6). Put

$$(5.1) \qquad a^b - 1 = p^e q(a) \qquad (p \nmid a),$$

where $(p-1)p^{e-1} \mid b$ and $q(a)$ is integral. Then it follows that

$$(5.2) \qquad \sum_{s=0}^{r} (-1)^{r-s} C_s^r S'_{sb+c}(p^h) = p^{re} \sum_{a=1}^{p^h}{}' a^c q(a), \ p \nmid a,$$

where $c = (p-1)u > 0$,

$$(5.3) \qquad S'_m(p^h) = \sum_{a=1}^{p^h} a^m, \ p \nmid a, \qquad S_m(p^h) = \sum_{a=1}^{p^h-1} a^m,$$

and h $(\geqq 1)$ will be fixed later. It is clear from (5. 3) that

$$(5.4) \qquad S'_{sb+c}(p^h) \equiv S_{sb+c}(p^h) \pmod{p^c}.$$

For $S_m(p^h)$ we shall require the well-known formula

$$(5.5) \qquad S_m(p^h) = (m+1)^{-1} \sum_{t=1}^{m+1} C_t^{m+1} B_{m+1-t} p^{th},$$

of which (3. 4) is the special case $h = 1$. If we rewrite (5. 5) in the form

$$(5.6) \qquad S_m(p^h) = \sum_{t=0}^{m} (t+1)^{-1} C_t^m B_{m-t} p^{(t+1)h},$$

we see that for $t \geqq p-1$, the t-th term in the right member of (5. 6) is divisible by at least p^{ph-2}. Thus in view of (5. 4) we get

$$(5.7)\qquad S'_{sb+c}(p^h) \equiv \sum_{t=0}^{p-2} (t+1)^{-1} C_t^{sb+c} B_{sb+c-t} p^{(t+1)h} \pmod{p^c, p^{ph-2}}.$$

Substitution in (5.2) leads to

$$(5.8)\qquad p^h \sum_{s=0}^{r} (-1)^{r-s} C_s^r B_m + \sum_{t=1}^{p-2} (t+1)^{-1} p^{(t+1)h} \sum_{s=0}^{r} (-1)^{r-s} C_s^r C_t^m B_{m-t}$$
$$\equiv 0 \pmod{p^c, p^{re}, p^{ph-2}}, \text{ where } m = sb + c.$$

Let us consider first the case $r < p$ and take $h = e$. Then it is easily seen, using (2.4), that provided $c > re$, the double sum in the left member of (5.8) is divisible by p^{re}; thus (5.8) becomes

$$(5.9)\qquad B^c(B^b - 1)^r \equiv 0 \pmod{p^{c-e}, p^{(r-1)e}, p^{(p-1)e-2}}.$$

Thus for $c > re$, the modulus is evidently $p^{(r-1)e}$ except perhaps when $r = p - 1$, $e = 1$, when the modulus is p^{r-2}.

Next let $r \geqq p$. We first remark that the modulus p^{ph-2} in (5.8) can be improved to p^{ph-1}; this is seen by examining the term corresponding to $t = p - 1$, namely,

$$(5.10)\qquad p^{ph-1} \sum_{s=0}^{r} (-1)^{r-s} C_s^r C_{p-1}^m B_{m-p+1}.$$

By the Staudt-Clausen theorem the sum in (5.10) is integral $(\text{mod } p)$ for $r \geqq p$. In the next place we verify that for $h \geqq e$, the double sum in (5.8) is divisible by $p^{h+(r-1)e}$, and thus we get

$$p^h B^c (B^b - 1)^r \equiv 0 \pmod{p^c, p^{re}, p^{h+(r-1)e}, p^{ph-1}}.$$

To find the most favorable value of h, it is only necessary to satisfy the inequality $ph - 1 \geqq re$, that is, $h \geqq (re+1)/p$. If therefore we choose h as the least integer $\geqq (re+1)/p$ (and thus $> e$) we conclude that

$$(5.11)\qquad B^c(B^b - 1)^r \equiv 0 \pmod{p^{re-h}} \qquad (r \geqq p).$$

Combining (5.9) and (5.11) we get

THEOREM 4. *Let* $(p-1)p^{e-1} \mid b$, $c = (p-1)u > re$; *then for* $r < p$ *we have*

$$(5.12)\qquad B^c(B^b - 1)^r \equiv 0 \pmod{p^{(r-1)e}},$$

except perhaps when $r = p - 1$, $e = 1$, *when the modulus is* p^{p-3}. *For* $r \geqq p$ *define* h *as the least integer* $\geqq (re+1)/p$; *then*

$$(5.13)\qquad B^c(B^b - 1)^r \equiv 0 \pmod{p^{re-h}}.$$

It is perhaps of interest to note what Theorem 4 becomes in the case $e = 1$. We state

THEOREM 4′. *Let* $p - 1 \mid b$, $c = (p-1)u > r$; *then* $B^c(B^b - 1)^r \equiv 0$ *holds* $\bmod p^{r-1}$, $\bmod p^{p-3}$ *or* $\bmod p^{r-h}$ *according as* $r < p-1$, $r = p-1$ *or* $r \geqq p$, *where* h *is the least integer* $\geqq (r+1)/p$.

Theorem 4′ evidently includes Theorem 2.

6. Proof of (1.7). Returning to (5.1) we raise both members to the $(r+1)$-th power and get after a little manipulation

$$(6.1)\qquad \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r (s+1)^{-1} (S'_{(s+1)b}(p^h) + p^{h-1} - p^h) = (r+1)^{-1} p^{e(r+1)} \sum_{a=1}^{p^h} q(a), \quad p \nmid a,$$

where $S'_m(p^h)$ has the same meaning as before, and $(p-1)p^{e-1} \mid b$. In view of (5.4), (6.1) becomes

$$(6.2)\qquad \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r (s+1)^{-1} (S_m(p^h) + p^{h-1} - p^h) \equiv (r+1)^{-1} p^{e(r+1)} \sum_{a} q(a) \pmod{p^b},$$

where for brevity we put $m = (s+1)b$. The argument is now similar to that of § 5. However, it is convenient to make some slight changes. In the first place we have

$$(s+1)^{-1}(S_m(p^h) + p^{h-1} - p^h) \equiv p^h(s+1)^{-1}(B_m + p^{-1} - 1) + b \sum_{t=1}^{p-1} t^{-1}(t+1)^{-1} p^{h(t+1)} C_{t-1}{}^{m-1} B_{m-t} \pmod{p^{h(p+1)+e-2}};$$

substitution in (6.2) gives

$$(6.3)\qquad p^h \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r (s+1)^{-1} (B_m + p^{-1} - 1) + b \sum_{t=1}^{p-1} t^{-1}(t+1)^{-1} p^{h(t+1)} \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r C_{t-1}{}^{m-1} B_{m-t} \equiv 0 \pmod{p^b, (r+1)^{-1} p^{e(r+1)}, p^{h(p+1)+e-2}}.$$

We now examine the double sum in (6.3). Take first $r < p$ and let $h = e$. We find that

$$(6.4)\qquad p^{h(t+1)} \sum_{s=0}^{r} (-1)^{r-s} C_s{}^r C_{t-1}{}^{m-1} B_{m-t} \equiv 0 \pmod{p^{(r+1)e}, p^b}$$

for $1 \leqq t \leqq p-1$. (The case $r=p-1$, $t=p-1$ requires special discussion as before.) Thus (6.3) implies

$$(6.5) \qquad p^h \sum_{s=0}^{r} (-1)^{r-s} C_s^r (s+1)^{-1}(B_m + p^{-1} - 1) \equiv 0 \pmod{M},$$

where

$$(6.6) \qquad M = (p^b, (r+1)^{-1}p^{e(r+1)}, p^{e(p+1)+e-2}, p^{(r+1)e+e-1}) = (p^b, (r+1)^{-1}p^{e(r+1)}).$$

In the next place let $r \geqq p$, $h \geqq e$. Then we can assert (6.4) with the modulus (p^{h+re}, p^b), also the last exponent in the modulus of (6.3) can be improved to $h(p+1)+e-1$. Hence a congruence of the form (6.5) holds with

$$M = (p^b, (r+1)^{-1}p^{e(r+1)}, p^{h(p+1)+e-1}, p^{h+re+e-1}).$$

We choose h so that $h(p+1)+e-1 \geqq h+re+e-1$, that is, $hp \geqq re$, and we find that again

$$(6.7) \qquad M = (p^b, (r+1)^{-1}p^{e(r+1)}).$$

Combining the several cases we get

THEOREM 5. *Let $(p-1)p^{e-1} \mid b$, $r \geqq 1$; then*

$$(6.8) \qquad \sum_{s=0}^{r} (-1)^{r-s} C_s^r (s+1)^{-1}(B_m + p^{-1} - 1) \equiv 0 \pmod{p^{b-h}, (r+1)^{-1}p^{e(r+1)-h}},$$

where $m=(s+1)b$, and $h=e$ for $r<p$, while h is the least integer $\geqq re/p$ for $r>p$.

Here again it may be of interest to restate the theorem in the special case $e=1$. We have

THEOREM 5′. *Let $p-1 \mid b$, $r \geqq 1$; then*

$$\sum_{s=0}^{r} (-1)^{r-s} C_s^r \frac{B_m + p^{-1} - 1}{s+1} \equiv 0 \pmod{p^{b-h}, (r+1)^{-1}p^{r+1-h}},$$

where $m=(s+1)b$ and h is the least integer $\geqq r/p$.

Generalizing (3.8), let us put

$$(6.9) \qquad \sigma_s = s^{-1}(B_{sb} + p^{-1} - 1);$$

then as in § 3 the left member of (6.8) can be denoted by $\Delta^r\sigma_1$, and we have

$$\Delta^r\sigma_a = \sum_{i=0}^{a-1} C_i^{a-1}\Delta^{r+1}\sigma_i \qquad (a \geqq 1). \tag{6.10}$$

By means of (6.10), Theorem 5 can be extended considerably. Indeed if M_r denotes the modulus in (6.8) then it follows immediately that

$$\Delta^r\sigma_a \equiv 0 \quad (M_r, \cdots, M_{r+a-1}). \tag{6.11}$$

The simplest case is that in which $a + r \leqq p - 1$, for in that case the denominators may be ignored; more generally the same remark applies when $r \equiv r_0 \pmod p$ and $a \leqq a + r_0 < p - 1$. Since $e(r+1) - h$ is a non-decreasing function of r it follows that the modulus in (6.11) is now

$$M = (p^{b-h_a}, p^{e(r+1)-h_1}) \qquad (a + r_0 \leqq p - 1), \tag{6.12}$$

where h_a is the least integer $\geqq (r + a - 1)e/p$ for $r + a - 1 \geqq p$; for $a + r \leqq p - 1$, (6.12) reduces to simply

$$M = (p^{b-e}, p^{er}) \qquad (a + r \leqq p - 1). \tag{6.13}$$

For the general case, however, we have

$$M = (p^{b-h_a}, (r+1)^{-1}p^{e(r+1)-h_1}, \cdots, (r+a)^{-1}p^{e(r+1)-h_a}), \tag{6.14}$$

and again h_a is the least integer $\geqq (r + a - 1)e/p$ ($h_a = e$ for $r + a \leqq p$). We have therefore

THEOREM 6. *Let* $(p-1)p^{e-1} \mid b$, $a \geqq 1$, $r \geqq 1$; *then* $\sigma^a(\sigma^b - 1)^r \equiv 0$ $(\mathrm{mod}\, M)$, *where* σ_a *is defined by* (6.9) *and* M *is determined by* (6.12), (6.13), (6.14).

In particular, Theorem 6 implies

THEOREM 7. *Let* $(p-1)p^{e-1} \mid b$, $a \geqq 1$, $1 \leqq r < p - 1$; *then*

$$\sigma^a(\sigma^b - 1)^r \equiv 0 \quad (\mathrm{mod}\, p^{b-h_a}, p^{er}), \tag{6.15}$$

where h_a *is the least integer* $\geqq (r + a - 1)e/p$.

For $a + r \leqq p - 1$, there is nothing to prove. For $a + r \geqq p$, it is only necessary to verify that $(r + a)^{-1}p^{ea-h_a}$, where $a \geqq 1$, is integral $(\mathrm{mod}\, p)$.

DUKE UNIVERSITY.

REFERENCES.

[1] G. Frobenius, "Über die Bernoullischen Zahlen und die Eulerschen Polynome," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1910), pp. 809-847.

[2] E. E. Kummer, "Über eine allgemeine Eigenschaft der rational Entwickelungs-coëfficienten einer bestimmten Gattung analytischer Functionen," *Journal für die reine und angewandte Mathematik*, vol. 41 (1851), pp. 368-372.

[3] Emma Lehmer, "On congruences involving Bernoulli numbers and the quotients of Fermat and Wilson," *Annals of Mathematics* (2), vol. 39 (1938), pp. 350-360.

[4] Niels Nielsen, *Traité élémentaire des nombres de Bernoulli*, Paris, 1923.

[5] H. S. Vandiver, "Certain congruences involving the Bernoulli numbers," *Duke Mathematical Journal*, vol. 5 (1939), pp. 548-551.

ON THE DERIVATIVES OF SOLUTIONS OF LINEAR, SECOND ORDER, ORDINARY DIFFERENTIAL EQUATIONS.*

By Philip Hartman.

In the differential equation

$$(1) \qquad y'' + qy = 0, \qquad (' = d/dt),$$

let $q = q(t)$ be continuous for $0 \leqq t < \infty$. The object of this note is to obtain estimates for the first derivative y' of a solution in terms of estimates for the solution y itself. Applications of these results to the problem of the location of points of the spectrum of boundary value problems associated with (1) can be made as in [5].

Theorem 1. *Let $Q = Q(t) > 0$ be a non-decreasing function satisfying*

$$(2) \qquad q \leqq Q.$$

Let $y = y(t)$ be a solution of (1) *and let*

$$(3) \qquad m(t) = \max |y(u)| \text{ for } |t - u| \leqq 2/Q^{\frac{1}{2}}(t + 2/Q^{\frac{1}{2}}(t)).$$

Then, for $t \geqq 2/Q^{\frac{1}{2}}(0)$,

$$(4) \qquad |y'(t)| \leqq 2Q^{\frac{1}{2}}(t + 2/Q^{\frac{1}{2}}(t))m(t).$$

For example, if, as $t \to \infty$, $y(t) = o(M(t))$, where $M(t)$ (> 0) is monotone, then $y'(t) = o(Q^{\frac{1}{2}}(t + \epsilon)M(t \pm \epsilon))$ for every fixed $\epsilon > 0$, where the choice of $\pm$ depends on whether M is non-decreasing or non-increasing. In this statement, o can be replaced by O. Thus, if y is $o(1)$ or $O(1)$, then y' is $o(Q^{\frac{1}{2}}(t + \epsilon))$ or $O(Q^{\frac{1}{2}}(t + \epsilon))$, as $t \to \infty$, for every fixed $\epsilon > 0$. (The occurrence of an arbitrary $\epsilon > 0$ is permissible because, in these o- and O-statements, there is no loss of generality in replacing Q by const. Q and supposing, therefore, that $2/Q^{\frac{1}{2}} \leqq \epsilon$.)

Proof. The proof depends on a simple Tauberian argument used in [2] to show that if

$$(5) \qquad q \leqq C, \qquad (C = \text{const.} > 0),$$

* Received July 26, 1952.

then $y = o(1)$ implies $y' = o(1)$. It was shown in [3], p. 650, that this argument implies the following: If (5) holds on $0 \leqq t < \infty$ and if $t \geqq 2/C^{\frac{1}{2}}$, then there exists a $t^* = t^*(t)$ with the properties that $| t^* - t | \leqq 2/C^{\frac{1}{2}}$ and

$$|y'(t)| \leqq 2C^{\frac{1}{2}} |y(t^*)|. \tag{4 bis}$$

(Of course, t^* depends on the solution y).

In the proof of (4 bis), only the interval with the end points t and t^* is involved. Hence the inequality (5) is needed only on this interval; for example, it is sufficient that (5) hold for all non-negative t-values not exceeding $t + 2/C^{\frac{1}{2}}$.

Theorem 1 will, therefore, follow if it is shown that (5) holds on the interval $[0, t + 2/C^{\frac{1}{2}}]$, where $C = Q(t + 2/Q^{\frac{1}{2}}(t))$. In view of (2) and the fact that $dQ \geqq 0$, it is sufficient to verify that

$$Q(t + 2/C^{\frac{1}{2}}) \leqq C \quad [= Q(t + 2/Q^{\frac{1}{2}}(t))].$$

But this inequality follows from the non-decreasing property of Q, since $C \geqq Q(t)$.

Theorem 1 and a result of Sears [7] lead to

THEOREM 2. *Let q, Q satisfy the conditions of Theorem 1. In addition, let*

$$\int^{\infty} dt/Q(t) = \infty; \tag{6}$$

for example, let

$$\liminf_{t \to \infty} Q(t)/t < \infty. \tag{7}$$

Let (1) possess a solution $y = y(t)$ ($\not\equiv 0$) which is bounded as $t \to \infty$. Then no solution of (1) linearly independent of $y(t)$ is of class $L_2(0, \infty)$. In particular, if all solutions of (1) are bounded (for example, if q is positive and non-decreasing), then no solution is of class $L_2(0, \infty)$.

The case $Q = \text{const.}$ of this assertion is known, as is the last parenthetical part concerning a positive, non-decreasing q; see [5], p. 214 and [4], p. 303, respectively.

It will be clear from the proof that if the assumptions (6) on Q and $y = O(1)$ on y are changed to

$$\int^{\infty} dt/Q(t)M^2(t) = \infty \text{ and } y = O(M(t)),$$

respectively, where M is positive and non-decreasing, then the first assertion of Theorem 2 remains valid and the second has an appropriate analogue.

Proof. Suppose that Theorem 2 is false; so that (1) has a solution $y = x(t)$ which is of class $L_2(0,\infty)$ and which is linearly independent of $y = y(t)$. Then (2) and $dQ \geqq 0$ imply that $x'(t)/Q^{\frac{1}{2}}(2t)$ is of class $L_2(0,\infty)$; [7]. Hence $x'(t)y(t)/Q^{\frac{1}{2}}(2t)$ is of class $L_2(0,\infty)$, since $y(t)$ is bounded. (Actually, $Q(2t)$ can be replaced by $Q(t+\epsilon)$.) On the other hand, Theorem 1 shows that $y'(t)/Q^{\frac{1}{2}}(2t)$ is bounded. Hence $x(t)y'(t)/Q^{\frac{1}{2}}(2t)$ is of class $L_2(0,\infty)$.

Since $x(t)$ can be multiplied by a constant factor ($\neq 0$), it can be supposed that the Wronskian of x and y is 1, $x'y - xy' \equiv 1$. If this relation is divided by $Q^{\frac{1}{2}}(2t)$, it follows from the linearity of the L_2-space, that $Q^{-\frac{1}{2}}(2t)$ is of class $L_2(0,\infty)$. But this contradicts (6) and implies Theorem 2.

THEOREM 3. *Let* $q^+ = q^+(t) = \max(0, q(t))$. *Let*

$$\int_0^t q^+(s)\,ds = O(t^2), \text{ as } t \to \infty. \tag{8}$$

Let (1) *possess a solution* $y = y(t)$ *which is bounded as* $t \to \infty$. *Then the assertions of Theorem* 2 *are valid.*

The relationship between Theorems 2 and 3 is clear. In Theorem 2, $Q(t)$ can be $O(t)$ (but not $O(t^{1+\epsilon})$). The inequality (8) is implied by $q \leqq \text{const. } t$, but not conversely.

As in Theorem 2, the assumptions (8) on q and $y = O(1)$ on y can be modified to

$$\int_0^t q^+M^2 ds = O(t^2) \text{ and } y = O(M(t)),$$

respectively, where M is a continuous function satisfying $0 \leqq M(t) = O(t^{3/2})$, as $t \to \infty$.

Theorem 3 implies the corresponding results of Fischel [1], where, instead of (8), it is assumed that

(i) $q > \text{const.} > -\infty$ and $q(t_2) - q(t_1) \leqq \text{const.}\,(t_2 - t_1)$ for $0 \leqq t_1 < t_2 < \infty$

or that

(ii) $|q(t_2) - q(t_1)| \leqq \text{const.}\,|t_2 - t_1|$

or that

(iii) $q(t)$ — const. is of class $L_2(0,\infty)$.

Of course, the cases (i) and (ii) are also implied by Theorem 2. In [4], p. 298, it was shown that (i), without the assumption $q >$ const., implies that if y is a solution of (1) of class $L_2(0,\infty)$, then $y' = O(1)$. It was shown in [1] that, in the case (iii), if $y = O(1)$, then $y' = O(1)$. It is easily seen that this proof in [1] is valid if $q(t)$ — const. is of class $L_p(0,\infty)$ for some $p \geqq 1$ or, more generally, if

$$\int_t^{t+1} |q(s)|\, ds = O(1), \text{ as } t \to \infty. \tag{9}$$

[It may be mentioned that Fischel's proof for the cases (i) and (ii) is incorrect (cf., p. 178, (α), (II)) in that he concludes (9) from the inequality $|q(t)| \leqq \text{const.}\, t$. In fact, it is not true that, in these cases, $y = O(1)$ implies $y' = O(1)$; this follows readily from standard asymptotic formulae for the case $q = t$.]

Proof of Theorem 3. According to [6], p. 150, if (1) has a solution $y = y(t)\ (\not\equiv 0)$ satisfying

$$\int_0^t y'^2 ds = O(t^2),$$

then no solution linearly independent of y is of class $L_2(0,\infty)$. (There is no assumption concerning q or y itself.) Hence, in view of (8), it is sufficient to show that if (1) has a solution y which is $O(1)$, as $t \to \infty$, then

$$\int_0^t y'^2 ds = O(1 + \int_0^{2t} q^+ ds). \tag{10}$$

To this end, let (1) be multiplied by y. If it is noted that

$$yy'' = (yy')' - y'^2$$

and that $qy^2 \leqq q^+y^2$, there results the inequality

$$y'^2 \leqq (yy')' + q^+y^2;$$

cf. [8], where this device is used for a similar purpose. The assumption $y = O(1)$ and a quadrature give

$$\int_0^t y'^2 ds \leqq yy' + O(1 + \int_0^t q^+ ds).$$

If t is sufficiently large, the interval $[t, 2t]$ contains a point u at which

$y(u)y'(u) \leqq 1$. For otherwise $\frac{1}{2}(y^2(2t) - y^2(t)) > t$ which, for large t, contradicts $y = O(1)$. Hence

$$\int_0^u y'^2 ds \leqq 1 + O(1 + \int_0^u q^+ ds).$$

Since $t \leqq u \leqq 2t$, the integral on the left (right) is not increased (decreased) if u is replaced by t ($2t$). This proves (10) and Theorem 3.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] B. Fischel, "The continuous spectra of certain differential equations," *Journal of the London Mathematical Society*, vol. 27 (1952), pp. 175-179.

[2] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[3] ———, "On bounded Green's kernels for second order linear ordinary differential equations," *American Journal of Mathematics*, vol. 73 (1951), pp. 646-656.

[4] ——— and A. Wintner, "Criteria of non-degeneracy for the wave equation," *ibid.*, vol. 70 (1948), pp. 295-308.

[5] ——— and A. Wintner, "On the location of spectra of wave equations," *ibid.*, vol. 71 (1949), pp. 214-217.

[6] ——— and A. Wintner, "On the derivatives of the solutions of one-dimensional wave equations," *ibid.*, vol. 72 (1950), pp. 148-156.

[7] D. B. Sears, "Note on the uniqueness of the Green's functions associated with certain differential equations," *Canadian Journal of Mathematics*, vol. 2 (1950), pp. 314-325.

[8] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

THE CRITICAL SET OF A CONVEX BODY.*

By V. L. Klee, Jr.**

1. Introduction. Suppose C is a convex body in E^n and $x \varepsilon \operatorname{Int} C$. For each $y \varepsilon FC$ let $\rho_C(y, x) = yx/yz$, where $[y, z]$ is the chord containing $[y, x]$. Let $\rho_C(x) = \sup_{y \varepsilon FC} \rho_C(y, x)$ and $S_C(x) = \{y \mid y \varepsilon FC \text{ and } \rho_C(y, x) = \rho_C(x)\}$. The number $\inf_{x \varepsilon \operatorname{Int} C} \rho_C(x)$ will be called the *critical ratio* of C and denoted by r_C. The set $\{x \mid \rho_C(x) = r_C\}$ will be called the *critical set* of C and denoted by C^*. These concepts were investigated in E^2 by B. H. Neumann [6] and in E^n by Süss [7], Hammer [1, 2, 3], and Hammer and Sobczyk [4]. (I am indebted to Dr. Hammer for supplying me with manuscripts of [3] and [4].) The present paper provides a more detailed discussion in E^n.

Neumann [6] proved that if $C \subset E^2$, then $\frac{1}{2} \leqq r_C \leqq \frac{2}{3}$ (with $r_C = \frac{2}{3}$ only for a triangle), C^* is a single point x, and $S_C(x)$ contains at least three points. For $C \varepsilon E^n$, Süss [7] and Hammer [1] showed that $\frac{1}{2} \leqq r_C \leqq n/(n+1)$, Hammer [3] that C^* is convex, and Hammer and Sobczyk [4] that C^* need not be degenerate. We prove here (3.9) that if $C \subset E^n$, then $r_C(1 - r_C)^{-1} + \dim C^* \leqq n$ (and that if $\geqq n - 1$, then for no point $x \varepsilon C^*$ is $S_C(x)$ contained in an open half-space parallel to C^* but not containing C^*). This inequality implies $\dim C^* \leqq n - 2$. Furthermore, for each $x \varepsilon C^*$, $S_C(x)$ contains at least $(1 - r_C)^{-1}$ points. By using a result of Hammer [1], we show (4.6) that if $C \subset E^n$, then $r_C = n/(n+1)$ if and only if C is an n-simplex. This was stated without proof by Süss [7].

Section 5 contains some extensions of Helly' theorem [5] on the intersection of convex sets, and is related to the rest of the paper only in that (5.5) is used to prove (3.8).

2. Notation and terminology. The interior, closure, boundary, ϵ-neighborhood, dimension, and convex hull of a set X will be denoted by $\operatorname{Int} X$, $\bar{X}$, FX, $N_\epsilon X$, $\dim X$, and $\operatorname{conv} X$ respectively. Set-theoretic union, intersection, and relative difference are indicated by $\cup$, $\cap$, and $\sim$, $+$ and $-$ being reserved for vector addition and subtraction. The intersection of all sets in a family

* Received February 27, 1952; revised September 22, 1952.

** National Research Fellow.

Q of sets is denoted by πQ. E^n is Euclidean n-space, O its origin, and Λ the empty set. $\{x \mid P(x)\}$ is the set of all x for which $P(x)$ is true. If x and y are points of E^n, then xy denotes the distances between them, $[x, y]$ the segment joining them, and $\overrightarrow{xy}$ the ray which emanates from x and passes through y. (If $xy = 0$, then $[x, y] = \overrightarrow{xy} = \{x\}$.) We use *convex body* to mean a bounded closed convex set whose interior is not empty. *C will always represent a convex body in E^n.* If K is a convex set in E^n, $\operatorname{lm} K$ will denote the smallest linear manifold containing K and $E(K)$ the interior of K relative to $\operatorname{lm} K$. Points of $E(K)$ will be called *equilibrium points* of K. "If and only if" will be rendered by "iff."

3. The basic inequality. In this section we establish the basic inequality stated in the introduction.

(3.1) *Suppose $x \in C^*$, $Z = \{z \mid z = \overrightarrow{yx} \cap FC$ for some $y \in S_C(x)\}$, and $\mathfrak{G}$ is the set of all half-spaces which intersect Z but not* $\operatorname{Int} C$. *Then $\pi\mathfrak{G}$ is empty.*

Proof. We suppose without loss of generality that $x = O$. If $G \in \mathfrak{G}$, then since $O \notin G$, $tG \subset \operatorname{Int} G$ for each $t > 1$. Hence if $\pi\mathfrak{G}$ is non-empty there is a point $p \in \pi\{\operatorname{Int} G \mid G \in \mathfrak{G}\}$. An easy argument using the compactness of Z and the upper semi-continuity of the supporting cone on FC then shows that there is an $\epsilon > 0$ such that if $Z' = FC \cap \overline{N_\epsilon Z}$ and $\mathfrak{G}'$ is the family of all half-spaces intersecting Z' but not $\operatorname{Int} C$, then $p \in \pi\mathfrak{G}'$.

Now consider an arbitrary $z \in Z'$. If $z + tp \in \operatorname{Int} C$ for some $t > 0$, then $(1 + t)^{-1}(z + tp) = q \in \operatorname{Int} C$, which is impossible since for some G we have $q \in [z, p] \subset G \in \mathfrak{G}'$. Thus $z + [0, \infty)p$ misses $\operatorname{Int} C$. If $z + (-\infty, 0)p$ also misses $\operatorname{Int} C$, then by the separation theorem for convex sets there is a $G \in \mathfrak{G}'$ such that $z + (-\infty, \infty)p \subset FG$. But then $z - p \in G$ and $p \in G$, whence $z \in G + G \subset \operatorname{Int} G$, a contradiction. Thus for each $z \in Z'$ there is a $t_z > 0$ such that $z - t_z p \in \operatorname{Int} C$. A simple argument using the compactness of Z' shows there is an $\eta > 0$ such that $z - \eta p \in \operatorname{Int} C$ whenever $z \in Z'$. Let $Y = \{y \mid y = \overrightarrow{zO} \cap FC$ for some $z \in Z'\}$. Then for each positive $\epsilon < \eta \inf\{\rho_C(y, O) \mid y \in Y\}$ there is a $\delta_\epsilon > 0$ such that $\rho_C(y, -\epsilon p) < r_C - \delta_\epsilon$ whenever $y \in Y$. And since $\rho(u, O) \mid u \in FC$ attains its maximum only on $S_C(O)$, there is for each $a > 0$ a $b(a) > 0$ such that $\rho_C(u, v) < r_C - b(a)$ whenever $u \in FC \sim N_a S_C(O)$ and $v \in N_{b(a)} O$. For sufficiently small a, $N_a S_C(O) \cap FC \subset Y$,

and for $\epsilon < b(a)/Op$ we have $\rho_C(u, -\epsilon p) < r_C - b(a)$ whenever $u \in FC \sim Y$ and $\rho_C(u, -\epsilon p) < r_C - \delta_\epsilon$ whenever $u \in Y$. Thus $\rho_C(-\epsilon p) < r_C$, a contradiction completing the proof.

(3.2) *Suppose $x \in \operatorname{Int} C$, $a \in S_C(x)$, $b = \overrightarrow{ax} \cap FC$, H_b is a hyperplane supporting C at b, and $H_a = H_b + (a - b)$. Then H_a supports C at a.*

Proof. Suppose FC contains a point q on the side of H_a away from b and let $q_a = \overrightarrow{xq} \cap H_a$, $q_b = \overrightarrow{qx} \cap H_b$, and $q' = \overrightarrow{qx} \cap FC$. Then

$$qx/qq' = qx/(qq_b - q'q_b) \geqq qx/qq_b$$
$$= (qq_a + q_a x)/(qq_a + q_a q_b) > q_a x/q_a q_b = ax/ab,$$

contradicting the hypothesis that $a \in S_C(x)$.

A similarly simple argument yields

(3.3) *Suppose x is an equilibrium point of C^* and $y \in S_C(x)$. Then $y + r_C^{-1}(C^* - y) \subset FC$ and $y \in S_C(p)$ for each $p \in C^*$.*

(3.3) shows that if C^* contains more than one point, FC must contain line segments. It shows also that the set $S_C(p)$ does not vary as p ranges over $E(C^*)$. We will henceforth denote this set by $C\dagger$.

(3.4) *Suppose O is an equilibrium point of C^*, L' a subspace of E^n complementary to* $\operatorname{lm} C^*$, *and $x' \mid x$ is the projection of E^n onto L' for which $C^{*\prime} = \{O\}$. Then $\rho_{C'}(O) = r_C$ and $S_{C'}(O) = S_C(O)$.*

Proof. Let $q \in FC'$. Then $q = y'$ for some $y \in FC$. By definition of r_C, $(1 - r_C)^{-1} r_C(-y) \in C$, so $(1 - r_C)^{-1} r_C(-q) \in C'$ and $\rho_{C'}(q, O) \leqq r_C$, with equality implying $y \in S_C(O)$. On the other hand, if $y \in S_C(O)$ and $z = \overrightarrow{yO} \cap FC$, it follows from (3.2) and (3.3) that C is supported at y and z by hyperplanes parallel to C^*. Hence $\rho_{C'}(O) \geqq r_C$ and $\rho_{C'}(y', O) = r_C$ whenever $y \in S_C(O)$. This completes the proof.

(3.5) *Suppose $Z \subset FC$, $O \in$* conv *Z, and $\rho_C(z, O) > m(m+1)^{-1}$ for each $z \in Z$. Then Z has a subset Z' such that* dim conv $Z' > m$ *and O is an equilibrium point of* conv Z'.

Proof. Let k be the smallest integer for which O belongs to a k-simplex having vertices in Z, and let $z_0, \cdots, z_k$ be such a set of vertices. Then

we have $O = \sum_{i=0}^{k} a_i z_i$ with each $a_i > 0$, so O is an equilibrium point of $\operatorname{conv}\{z_0, \cdots, z_k\}$. Let $-y_i = \overrightarrow{z_i O} \cap FC$. Then $y_0 = -\sum_{i=1}^{k} t_i y_i$, with each $t_i > 0$. For $1 \leqq j \leqq k$,

$$-y_j = (mt_j)^{-1}(my_0) + \sum_{1 \leqq i \leqq k,\, i \neq j} t_i (mt_j)^{-1}(my_i),$$

where each point $my_l (0 \leqq l \leqq k)$ is interior to C and the coefficients are all positive. Since $-y_j$ is not interior to C, the sum of the coefficients is not less than 1, and hence

$$mt_j \leqq 1 + \sum_{1 \leqq i \leqq k,\, i \neq j} t_i \qquad (1 \leqq j \leqq k).$$

Summing on j yields $m\sigma \leqq k + (k-1)\sigma$, with $\sigma = \sum_{i=1}^{k} t_i$, whence $k/(m+1) \geqq \sigma/(\sigma+1)$. But $\operatorname{conv}\{my_1, \cdots, my_k\} \subset \operatorname{Int} C$ and $-y_0 \in FC$, so $\sigma > m$. Using this and the previous inequality we obtain $k/(m+1) > m/(m+1)$, whence $k > m$ and the proof is complete.

(3.6) *Suppose C is not contained in any open half-space parallel to* $\operatorname{lm} C^*$ *but not containing it. Then* $r_C(1 - r\ )^{-1} + \dim C^* \leqq n$.

Proof. Suppose $O \in E(C^*)$ and let L' and C' be as in (3.4). From (3.4) we have $\rho_{C'}(z, O) = r_C$ for each $z \in S_C(O)'$, and the hypothesis above on $S_C(x)$ insures that $O \in \operatorname{conv} S_C(O)'$. Then (3.5) gives $\dim L' > m$ for each m such that $m(m+1)^{-1} < r_C$. Since $\dim L' = n - \dim C^*$, we thus have $n - \dim C^* > m$ whenever $m < r_C(1 - r_C)^{-1}$, whence $r_C(1 - r_C)^{-1} + \dim C^* \leqq n$.

(3.7) *Suppose B is a convex set in E^n, $0 < t < \mu = (1 - r_B)r_B^{-1}$, $A = \operatorname{conv}(B \cup -tB)$, $x \in B^* \sim \{O\}$, and $s > t(\mu + 1)(\mu - 1)^{-1}$. Then $-sx$ is in every half-space which intersects $-tB$ but not* $\operatorname{Int} A$.

Proof. Let $b \in B$ and the linear functional $f (\neq 0)$ be such that $f(a) \geqq f(-tb)$ whenever $a \in A$. Clearly $x - \mu(b - x) \in B$, so $(\mu + 1)f(x) - \mu f(b) \geqq f(-tb)$. From this it follows that

$$f(-x) \leqq t^{-1}(\mu + 1)^{-1}(\mu - t) f(-tb),$$

whence $f(-sx) < f(-tb)$ and the proof is complete.

(3.8) *Suppose C is such that $r_C(1 - r_C)^{-1} + \dim C^* \geqq n - 1$. Then*

$C\dagger$ is not contained in any open half-space which is parallel to C^* but doesn't contain C^*.

Proof. Simple continuity considerations show that it suffices to consider the case $\cdots > n-1$. Let $k = \dim C^* + 1$ and $m = r_C(1-r_C)^{-1}$. Then $m > n-k$. We assume without loss of generality that $O \in E(C^*)$. If the conclusion of (3.8) is false, there is an open half-space H such that $O \notin H$, $-C\dagger \subset H$, and FH is parallel to $\operatorname{lm} C^*$. Let Z and $\mathfrak{G}$ be as in (3.1). Then $\Gamma = \{\operatorname{Int} G \mid G \in \mathfrak{G}\}$ is an *0-closed family* (in the sense of § 5) of half-spaces, each containing a translate of $\operatorname{lm} C^*$. Since $Z = -m^{-1}C\dagger$ it follows from (3.7) that each $n-k$ members of Γ have common points in H arbitrarily far from FH, and then from (5.5) that $\pi\Gamma$ is non-empty, contradicting (3.1) and completing the proof.

Combining (3.7) and (3.8), we have

(3.9) *Suppose C is a convex body in E^n. Then $r_C(1-r_C)^{-1} + \dim C^* \leqq n$. If $r_C(1-r_C)^{-1} + \dim C^* \geqq n-1$, then*

$$E(C^*) \subset \operatorname{Int}\operatorname{conv}(C^* \cup C\dagger).$$

Directly from (3.9) we obtain

(3.10) *With C as in (3.9), $\dim C^* \leqq n-2$ and $r_C \leqq n/(n+1)$. If $r_C > (n-1)/n$, then C^* is a single point and $C^* \subset \operatorname{Int}\operatorname{conv} C\dagger$.*

(3.11) *$C\dagger$ contains at least $(1-r_C)^{-1}$ points.*

Proof. For $r_C > (n-1)/n$ it follows from (3.10) that $C\dagger$ contains at least $n+1$ points. So suppose $r_C \leqq (n-1)/n$ and let $B = \operatorname{conv} C\dagger$. Since $(1-r_C)^{-1} \leqq n$, if $C\dagger$ does not contain at least $(1-r_C)^{-1}$ points, B is a j-simplex for some $j \leqq n-1$ and $r_B = (j-1)/j$. Defining Z as in (3.1) (and taking $O \in E(C^*)$) we have $Z = -(1-r_C)r_C^{-1}C$. Thus if $(1-r_C)r_C^{-1} < 1/j$, the conclusion of (3.7) contradicts that of (3.1). Hence $(1-r_C)r_C^{-1} \geqq 1/j$, from which it follows that $j+1 \geqq (1-r_C)^{-1}$, completing the proof.

4. Characterization of the simplex. By a *conical convex body* in E^n we mean a set C of the form $C = \operatorname{conv}(\{p\} \cup K)$, where K is an $(n-1)$-dimensional convex body contained in a hyperplane H in E^n and p is a point of $E^n \sim H$. Each such point p will be called a *conical vertex* of C. In this

section we examine r_C, C^*, and $C\dagger$ for such bodies and obtain, in particular, a proof that in E^n, r_C attains its maximum only for the n-simplex. We use (4.1) and (4.2) below, which are given implicitly by Neumann [6].

(4.1) *For* $x \in C$, $\rho_C(x) = \inf\{t \mid x - tC \supset (1-t)C\}$.

Proof. Let $K = C - x$. Then $\rho_C(x) = \rho_K(O)$ and it can be verified that $x - tC$ contains $(1-t)C$ iff $-tK$ contains $(1-t)K$. Thus it suffices to consider only the case $x = O$. If $O \in FC$, the above expression gives $\rho_C(O) = 1$, which is correct. Now suppose $O \in \operatorname{Int} C$, $\rho_C(O) = \epsilon$, $y \in FC$, and $-a^{-1}y = \overrightarrow{yO} \cap FC$. Then $(\epsilon - 1)\epsilon^{-1}y = b(-a^{-1}y)$, with $b = (1-\epsilon)\epsilon^{-1}a$. However, it follows from the definition of ϵ that $a \leqq \epsilon(1-\epsilon)^{-1}$, whence $b \leqq 1$ and $(\epsilon - 1)\epsilon^{-1}y \in C$. Thus $(\epsilon-1)\epsilon^{-1}C \subset C$, or $-\epsilon C \supset (1-\epsilon)C$, and $\epsilon \leqq \inf\{t \mid \cdots\}$. But if $-a^{-1}y \in S_C(O)$, then $a = \epsilon(1-\epsilon)^{-1}$, so $b = 1$, and $(\epsilon-1)\epsilon^{-1}y \in FC$. This implies that $\epsilon \geqq \inf\{t \mid \cdots\}$ and shows that for $x \in \operatorname{Int} C$, $\rho_C(x)$ can be obtained as stated. A similar but simpler argument yields

(4.2) *For* $x \in C$, $S_C(x) = [(1-\rho_C(x))^{-1}(x - \rho_C(x))FC] \cap C$. *Also* $r_C = \inf\{t \mid (1-t)C$ *is contained in some translate of* $-tC\}$, *and*

$$C^* = \{x \mid x - r_C C \supset (1-r_C)C\}.$$

It is more convenient here to work with the numbers

$$\mu_C(x) = \rho_C(x)/(1-\rho_C(x)) \text{ and } m_C = \inf_{x \in \operatorname{Int} C} \mu_C(x).$$

From (4.1) and (4.2) it follows that

(4.3) $\mu_C(x) = \inf\{t \mid C \subset (1+t)x - tC\}$ *and*

$$S_C(x) = [(1+\mu_C(x))x - \mu_C(x)FC] \cap C.$$

(4.4) *Suppose K is an $(n-1)$-dimensional convex body contained in a hyperplane in* $E^n \sim \{O\}$, $0 \leqq a < 1$, *and* $C = \operatorname{conv}(aK \cup K)$. *Let* $t = (1+m_K)/(2+m_K-a)$ *and*

$$B = aK \cap [(1+m_K)\mathrm{K}^* - (1+m_K-a)K].$$

Then $m_C = 1 + m_K - a$, $C^* = tK^*$, *and* $C\dagger = K\dagger \cup B$. *If* $k \in K^*$, *then* $S_C(tk) = S_K(k) \cup B$.

Proof. (We actually prove only the assertions about m_C and C^*. The other proofs are not trivial but are omitted as being of less interest.) Note

that $C = [a, 1]K$, and consider an arbitrary $\lambda k \varepsilon \operatorname{Int} C$ (with $k \varepsilon \operatorname{Int} K$, $\lambda \varepsilon (a, 1)$). Then the following statements are equivalent:

(1) $r \geqq \mu_C(\lambda k)$; (2) $C \subset (1 + r)\lambda k - rC$;

$$(3) \quad \begin{cases} \text{(3a) there is a number } \alpha \varepsilon [a, 1] \text{ such that } K \subset (1 + r)\lambda k - r\alpha K; \\ \text{(3b) there is a number } \beta \varepsilon [a, 1] \text{ such that } aK \subset (1 + r)\lambda k - r\beta K. \end{cases}$$

$$(4) \quad \begin{cases} \text{(4a) } s \geqq ra \text{ and } K \subset (1 + s)k - sK, \text{ where } s = \lambda + r\lambda - 1; \\ \text{(4b) } \lambda + r\lambda - a \leqq r. \end{cases}$$

Equivalence of (1) and (2) follows from (4. 3), of (2) and (3) from the definition of C. Since K is contained in a hyperplane missing O, (3a) implies $\alpha = (\lambda + r\lambda - 1)/r$, from which the equivalence of (4a) and (3a) follows easily. Similarly, (3b) implies $\beta = (\lambda + r\lambda - a)/r$, which with $\beta \leqq 1$ implies (4b). On the other hand, if (4a) and (4b) are both valid then for $r = 0$, (3b) is trivial, and for $a > 0$, (3b) follows from the inclusion "$K \subset \cdots$" of (4a) and the fact that $(\lambda + r\lambda)/a - 1 > s$. Thus the equivalence of (1)-(4) is established, and $\mu_C(\lambda k)$ is the smallest number r for which (4) is valid.

Let $\mu = \mu_C(\lambda k)$ and $m = m_K$. The inclusion of (4a) implies $s \geqq \mu_K(k)$, so from $r = (1 + s - \lambda)/\lambda$ we get $\mu \geqq (1 + m - \lambda)/\lambda$, with equality implying $k \varepsilon K^*$. For $\lambda \leqq t$ this yields $\mu \geqq 1 + m - a$, with equality implying $\lambda = t$ and $k \varepsilon K^*$. From (4b) we get $r \geqq (\lambda - a)/(1 - \lambda)$, which for $\lambda < t$ implies $\mu < 1 + m - a$. We have thus shown that $\mu_C(\lambda k) \geqq 1 + m - a$, with equality implying $\lambda = t$ and $k \varepsilon K^*$. Furthermore, it can be verified that (4) is satisfied by $r = 1 + m - a$, $\lambda = t$, and $k \varepsilon K^*$, so the assertions of (4. 4) about m_C and C^* have been proved.

As a special case of (4. 4) we have

(4. 5) *Suppose K is an $(n - 1)$-dimensional convex body contained in a hyperplane in $E^n \sim \{O\}$, and $C = \operatorname{conv}(\{O\} \cup K)$. Then $r_C = 1/(2 - r_K)$, $C^* = r_C K^*$, and $C\dagger = K\dagger \cup \{O\}$. If $k \varepsilon K^*$, then $S_C(r_C k) = S_K(k) \cup \{O\}$.*

In [1], Hammer has proved (H): *Suppose C is a convex body in E^n and x is the centroid of C. Then $\rho_C(x) \leqq n/(n + 1)$ and equality implies that each point of $S_C(x)$ is a conical vertex of C.* From (H) and (4. 5) we have the following result, stated without proof by Süss [7; p. 127].

(4. 6) *If C is a convex body in E^n and $r_C = n/(n + 1)$, then C is an n-simplex.*

Proof. From (H) it follows that C is a cone over some $(n-1)$-dimensional convex set K, and (4.5) shows that $r_K = (2r_C - 1)/r_C = (n-1)/n$. Thus the obvious inductive argument yields a proof.

From (4.5) and (3.10) we have

(4.7) *Suppose C has k conical vertices, with $k \leqq n-2$. Then* $\dim C^* \leqq n - 2 - k$ *and* $r_C \geqq (k+1)/(k+2)$.

In particular, every conical convex body in E^3 has a unique critical point.

5. Some intersection properties of convex sets. Consider a family Γ of convex sets in E^n, each $n+1$ of which have a common point. Helly [5] proved that if Γ is finite or all its sets are compact, then there is a point in common to all sets of Γ. We show below (5.3) that the conclusion holds also if the family Γ is 0-*closed* (as defined later), with no assumption of finiteness or compactness being necessary in this case. This is used in proving (5.5), whose application in (3.8) provides the only connection between this section and the rest of the paper.

(5.1) *Let R_1, R_2, and R_3 mean (respectively) "intersects," "is contained in," and "contains." Suppose Γ is a family of convex sets in E^n, K is a convex set in E^n, and for each $n+1$ sets of Γ there is a translate of K which R_j all $n+1$ of them. Suppose further that at least one of the following holds:*

(i) *Γ is finite;*

(ii) *K and the sets of Γ are all bounded and closed;*

(iii) *$j = 2$, K is open, and the sets of Γ are all bounded.*

Then some translate of K R_j all the sets of Γ.

Proof. For each $C \varepsilon \Gamma$ let $C' = \{p \mid (K+p)R_jC\}$. Then each C' is convex, under (ii) or (iii) is bounded and closed, and the R_j-hypothesis relating K and Γ implies that each $n+1$ sets C' have a common point. Thus by Helly's theorem there is a point $x \varepsilon \pi\{C' \mid C \varepsilon \Gamma\}$, and $K + x$ is the desired translate of K.

For $j = 1$, (5.1) was proved by Vincensini [8] in a different way. We shall need the following lemma.

(5.2) *Suppose C_α is a convergent sequence of convex sets in E^n and $p \in \operatorname{Int} \lim C_\alpha$. Then there is an integer k such that $p \in \operatorname{Int} \bigcap_{i \geqq k} C_i$.*

Proof. Let $C = \lim C_\alpha$. We suppose without loss of generality that the C_i's are uniformly bounded. (For some $M < \infty$, $xO < M$ whenever x is in any C_i.) If for no k is $p = O$ interior to $\bigcap_{i \geqq k} C_i$, then C_α admits a consequence K_α such that $r_\alpha \to 0$, where r_i is the radius of the largest open sphere centered at O and contained in K_i. By the support theorem for convex bodies there is for each i a unit vector v_i such that $r_i = \sup \{v_i \cdot x \mid x \in K_i\}$. Let $n(\alpha)$ be such that $v_{n(\alpha)}$ converges (say to v) and set $\epsilon_i = v_{n(i)} v$. Then for each $x \in K_{n(i)}$ we have

$$v \cdot x = (v - v_{n(i)}) \cdot x + v_{n(i)} \cdot x \leqq M\epsilon_i + r_{n(i)},$$

where $\epsilon_\alpha \to 0$ and $r_{n(\alpha)} \to 0$. Thus $C \subset \liminf K_{n(\alpha)} \subset \{x \mid v \cdot x \leqq 0\}$, contradicting the fact that $O \in \operatorname{Int} C$.

A family Φ of sets in E^n will be called 0-*closed* iff every set in Φ is open and $\operatorname{Int} F \in \Phi$ whenever F is the limit of a convergent sequence of sets of Φ. We have

(5.3) *Suppose Γ is an 0-closed family of convex sets in E^n, each $n+1$ of which have a common point. Then* $\operatorname{Int} \pi\Gamma \neq \Lambda$.

Proof. In view of (5.1) (for $j = 2$) it suffices to prove the assertion (A): For some integer m, each $n+1$ members of Γ contain a common $1/m$-sphere which is contained in $S_m = \{x \mid xO < m\}$. Suppose (A) is false. Then for each m there are sets $C_m^1, \cdots, C_m^{n+1}$, of Γ such that $S_m \cap C_m^1 \cap \cdots \cap C_m^{n+1}$ contains no $1/m$-sphere, and such sets can be chosen so that C_α^h converges for each h; say $C_\alpha^h \to C_0^h$. For each i, let $D_i = \bigcap_{1 \leqq j \leqq h} C_i^j$. We will show that D_α admits a convergent sequence $K_\alpha \to K_0$ such that $\operatorname{Int} K_0 = D_0$, and will use this fact to complete the proof of (5.3).

Notice that for every subsequence K_α of D_α,

$$\limsup K_\alpha \subset \limsup D_\alpha \subset D_0,$$

so it suffices to choose a convergent K_α such that $\operatorname{Int} D_0 \subset \operatorname{Int} \lim K_\alpha$. Let G_α be an arbitrary convergent subsequence of D_α. If $\operatorname{Int} D_0 = \Lambda$, let $K_\alpha = G_\alpha$. Otherwise, for each $p \in \operatorname{Int} D_0$ there is (by (5.2)) an integer $k(p)$

such that $p \in \operatorname{Int} \bigcap_{i \geqq i(p)}^{\infty} G_i$. By the Lindelöf theorem there is a sequence $p_1, p_2, \cdots$, of points in $\operatorname{Int} D_0$ such that $\operatorname{Int} D_0$ is covered by the associated open sets. Let $K_i = G_{k(p_i)}$. Then K_α is the desired subsequence.

Now since D_0 is non-empty, K_0 has an interior point p, so by (5.2) there are an integer j and an $\epsilon > 0$ such that

$$N_\epsilon p \subset \bigcap_{i \geqq j} K_i.$$

For $m > \max(1/\epsilon, 2pO)$, this contradicts the fact that $S_m \cap D_m$ contains no $1/m$-sphere. The proof of (5.3) is complete.

(5.4) *Suppose* Γ *is an* 0-*closed family of convex sets in* E^n, $E^n \neq K \in \Gamma$, *and* $K \cap C_1 \cap \cdots \cap C_n$ *contains points at arbitrarily great distance from* FK *whenever* $\{C_1, \cdots, C_n\} \subset \Gamma$. *Then* $\operatorname{Int} \pi\Gamma \neq \Lambda$.

Proof. For $n = 1$, (5.4) is obvious, so suppose it has been proved for all $n < m$ and consider the case $n = m$. To show (in E^m) that $\operatorname{Int} \pi\Gamma \neq \Lambda$ it suffices, in view of (5.3), to show that if $\{C_0, C_1, \cdots, C_m\} \subset \Gamma$, then $\bigcap_{0 \leqq i \leqq m} C_i \neq \Lambda$. Suppose this is not the case and let $S = \bigcap_{1 \leqq i \leqq m} C_i$. Then C_0 and S are disjoint open convex sets, so can be separated by a hyperplane H disjoint from both. For K to contain H it is necessary that FK is either a translate of H or the union of two such translates. Since both $S \cap K$ and $C \cap K$ contain points arbitrarily far from FK, this is not possible, and thus $K \cap H$ is an open convex proper subset K' of H. For $1 \leqq i \leqq m$, let $C_i' = C_i \cap H$, and consider an arbitrary $m - 1$ sets C_i', say $\{C_i' \mid i \neq k\}$. Since $\bigcap_{i \neq k} C_i$ contains S and intersects C_0, it follows that $(\bigcap_{i \neq k} C_i') \cap K'$ contains points arbitrarily far from FK' (boundary relative to H), so by the inductive hypothesis, $\bigcap_{1 \leqq i \leqq k} C_i' \neq \Lambda$. This contradicts the fact that $S \cap H = \Lambda$, and completes the proof.

We finally obtain an extension of (5.4) of a sort available for every Helly-type theorem in E^n. It is given here explicitly because it is needed to prove (3.8).

(5.5) *Suppose* L *is a* k-*dimensional linear manifold in* E^n, Γ *is an* 0-*closed family of convex sets in* E^n, *each containing a translate of* L, $E^n \neq K \in \Gamma$, *and* $K \cap C_1 \cap \cdots \cap C_{n-k}$ *contains points at arbitrarily great distance from* FK *whenever* $\{C_1, \cdots, C_{n-k}\} \subset \Gamma$. *Then* $\operatorname{Int} \pi\Gamma$ *contains a translate of* L.

Proof. Let L' be a linear manifold complementary to L and for each $C \varepsilon \Gamma$ let $C' = C \cap L'$. Let $\Gamma' = \{C' \mid C \varepsilon \Gamma\}$. It can be verified that the triple L', K', Γ' satisfies the conditions placed in (5.4) on E^n, K, Γ, and the proof is completed by applying (5.4).

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] P. C. Hammer, "The centroid of a convex body," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 522-525.

[2] ———, "Convex bodies associated with a convex body," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 781-793.

[3] ———, "Associated convex bodies," unpublished manuscript.

[4] ——— and A. Sobczyk, "The critical points of a convex body," Abstract 112, *Bulletin of the American Mathematical Society*, vol. 57 (1951), p. 127.

[5] E. Helly, "über Mengen konvexer Körper mit gemeinschaftlichen Punkten," *Jahresbericht der Deutschen Mathematiker Vereinigung*, vol. 32 (1923), pp. 175-176.

[6] B. H. Neumann, "On some affine invariants of closed convex regions," *Journal of the London Mathematical Society*, vol. 14 (1939), pp. 262-272.

[7] W. Süss, "über eine Affininvariant von Eibereichen," *Archiv der Mathematik*, vol. 1 (1948-1949), pp. 127-128.

[8] P. Vincensini, "Sur une extension d'un theoreme de M. J. Radon sur les ensembles de corps convexes," *Bulletin de la Société Mathématique de Frances*, vol. 67 (1939), pp. 115-119.

ACYCLIC MODELS.*

By Samuel Eilenberg and Saunders MacLane.

1. Introduction. There are a number of situations in algebraic topology where one establishes the existence of chain transformations and chain homotopies, dimension by dimension, using the fact that certain homology groups of a local character are zero. Most of the applications can be derived from well known theorems dealing with acyclic carriers. Other investigations (e. g. complexes with operators [2] and homology theory of multiplicative systems [3]) lead to similar proofs in situations no longer covered by theorems on acyclic carriers. The present paper formulates a general theorem, which seems to subsume all the situations of this type hitherto encountered. The theorem is formulated in the language of categories and functors [1].

As applications, we give proofs of the theorems of this type encountered in [2] and [3]. However, the most important application is a new theorem, establishing, by this method, the equivalence of the singular homology theories based respectively on simplexes and on cubes. This result was prompted by recent work of J. P. Serre and H. Cartan.

Another application is included in the paper of Eilenberg-Zilber [4] immediately following.

2. Main definitions and results. Let $\mathfrak{A} = (A, \alpha)$ be a category with objects A and maps α. We shall assume given a set $\mathfrak{M}$ of objects in $\mathfrak{A}$ (called *model objects*).

Let T be any covariant functor on the category $\mathfrak{A}$ with values in the category $\mathfrak{G}$ of abelian groups. We define a new functor $\tilde{T}$ on $\mathfrak{A}$ to $\mathfrak{G}$ as follows. For each object $A \varepsilon \mathfrak{A}$, the group $\tilde{T}(A)$ is the free abelian group generated by the symbols (ϕ, m) where $\phi: M \to A$ is a map (in $\mathfrak{A}$), $M \varepsilon \mathfrak{M}$ and $m \varepsilon T(M)$. If $\alpha: A \to B$ in $\mathfrak{A}$ then $\tilde{T}(\alpha)$ is defined by $\tilde{T}(\alpha)(\phi, m) = (\alpha\phi, m)$. In addition we define a natural transformation $\Phi: \tilde{T} \to T$ by setting $\Phi(A)(\phi, m) = T(\phi)m$.

Definition. The functor T is said to be *representable* if there is a natural transformation $\Psi: T \to \tilde{T}$ such that the composition $\Phi\Psi: T \to T$ is the identity. Ψ is called a *representation* of Φ.

* Received May 26, 1952.

Let $\partial \mathfrak{G}$ denote the category of chain complexes and chain transformations, and let K be a covariant functor on $\mathfrak{A}$ to $\partial \mathfrak{G}$. For each object $A \varepsilon \mathfrak{A}$, the functor K then determines a complex $K(A)$ composed of chain groups $K_q(A)$ and boundary homomorphisms $\partial_q : K_q(A) \to K_{q-1}(A)$, with $\partial_{q-1}\partial_q = 0$. The groups $K_q(A)$ yield a functor K_q on $\mathfrak{A}$ to $\mathfrak{G}$ and the boundary operators yield natural transformations $\partial_q : K_q \to K_{q-1}$ with $\partial_{q-1}\partial_q = 0$.

Let K and L be two (covariant) functors on $\mathfrak{A}$ to $\partial \mathfrak{G}$. A *map* $f : K \to L$ is a family of natural transformations $f_q : K_q \to L_q$ such that $\partial_q f_q = f_{q-1}\partial_q$. If f_q is defined and satisfies this equation only for $q \leqq n$, we say that f is a map $K \to L$ in dimensions $\leqq n$.

Let $f, g : K \to L$ be two maps. A *homotopy* $D : f \simeq g$ is a sequence of natural transformations $D_q : K_q \to L_{q+1}$ such that

$$\partial_{q+1} D_q + D_{q-1}\partial_q = g_q - f_q. \tag{2.1}$$

If the maps D_q are defined and satisfy (2.1) only for $q \leqq n$, we say that D is a homotopy in dimensions $\leqq n$.

THEOREM Ia. *Let K and L be covariant functors on $\mathfrak{A}$ with values in $\partial \mathfrak{G}$, and let $f : K \to L$ be a map in dimensions $< q$. If K_q is representable and if $H_{q-1}(L(M)) = 0$ for each model $M \varepsilon \mathfrak{M}$, then f admits an extension to a map $K \to L$ in dimensions $\leqq q$.*

Proof. The objective is to define a natural transformation $f_q : K_q \to L_q$ such that $\partial_q f_q = f_{q-1}\partial_q$.

For each $m \varepsilon K_q(M)$, $M \varepsilon \mathfrak{M}$, we have $f_{q-1}\partial_q m \varepsilon L_{q-1}(M)$. Since $\partial_{q-1} f_{q-1} \partial_q = f_{q-1}\partial_{q-1}\partial_q = 0$ it follows that $f_{q-1}\partial_q m$ is a $(q-1)$-dimensional cycle in $L(M)$. Since $H_{q-1}(L(M)) = 0$ we may choose a chain $d(m) \varepsilon L_q(M)$ with $\partial_q d(m) = f_{q-1}\partial_q m$.

We now consider the functor K_q associated with K_q and define a natural transformation $\Lambda : \tilde{K}_q \to L_q$ by setting $\Lambda(A)(\phi, m) = L_q(\phi) d(m)$. We have

$$\begin{aligned}\partial_q \Lambda(A)(\phi, m) &= \partial_q L_q(\phi) d(m) = L_{q-1}(\phi)\partial_q d(m) \\ &= L_{q-1}(\phi) f_{q-1}\partial_q m = f_{q-1}\partial_q K_q(\phi) m = f_{q-1}\partial_q \Phi(A)(\phi, m)\end{aligned}$$

where $\Phi : \tilde{K}_q \to K_q$ is defined as above. Thus $\partial_q \Lambda = f_{q-1}\partial_q \Phi$. Now, let $\Psi : K_q \to \tilde{K}_q$ be a representation of K_q, and let $f_q = \Lambda\Psi : K_q \to L_q$. Then $\partial_q f_q = \partial_q \Lambda \Psi = f_{q-1}\partial_q \Phi\Psi = f_{q-1}\partial_q$, as desired.

THEOREM Ib. *Let K and L be covariant functors on $\mathfrak{A}$ with values in $\partial \mathfrak{G}$, let $f, g : K \to L$ be maps, and let $D : f \simeq g$ be a homotopy in dimensions*

$< q$. *If K_q is representable and if $H_q(L(M)) = 0$ for each model $M \varepsilon \mathfrak{M}$, then D admits an extension to a homotopy $f \simeq g$ in dimensions $\leqq q$.*

Proof. The objective is to define a natural transformation $D_q : K_q \to L_{q+1}$ with $\partial_{q+1} D_q + D_{q-1}\partial_q = g_q - f_q$.

For each $m \varepsilon K_q(M)$, $M \varepsilon \mathfrak{M}$, we have $(g_q - f_q - D_{q-1}\partial_q) m \varepsilon L_q(M)$. Since

$$\begin{aligned}\partial(g_q - f_q - D_{q-1}\partial_q) &= \partial_q g_q - \partial_q f_q - (\partial_q D_{q-1})\partial_q \\ &= g_{q-1}\partial_q - f_{q-1}\partial_q - (g_{q-1} - f_{q-1} - D_{q-2}\partial_{q-1})\partial_q = 0,\end{aligned}$$

it follows that $(g_q - f_q - D_{q-1}\partial_q)m$ is a q-cycle in $L(M)$. Since $H_q(L(M)) = 0$ we may choose a chain $e(m) \varepsilon L_{q+1}(M)$ with $\partial_{q+1} e(m) = (g_q - f_q - D_{q-1}\partial_q)m$. We now define $\Gamma : \tilde{K}_q \to L_{q+1}$ by setting $\Gamma(A)(\phi, m) = L_{q+1}(\phi) e(m)$. We have

$$\begin{aligned}\partial_{q+1}\Gamma(A)(\phi, m) &= \partial_{q+1} L_{q+1}(\phi) e(m) = L_q(\phi)\partial_{q+1} e(m) \\ &= L_q(\phi)(g_q - f_q - D_{q-1}\partial_q) m = (g_q - f_q - D_{q-1}\partial_q) K_q(\phi) m \\ &= (g_q - f_q - D_{q-1}\partial_q)\Phi(A)(\phi, m).\end{aligned}$$

Thus $\partial_{q+1}\Gamma = (g_q - f_q - D_{q-1}\partial_q)\Phi$. Now, let $\Psi : K_q \to \tilde{K}_q$ be a representation of K_q, and let $D_q = \Gamma\Psi : K_q \to L_{q+1}$. Then

$$\partial_{q+1} D_q = (g_q - f_q - D_{q-1}\partial_q)\Phi\Psi = g_q - f_q - D_{q-1}\partial_q,$$

as desired.

Theorems Ia and Ib imply

THEOREM II. *Let K and L be covariant functors on $\mathfrak{A}$ with values in $\partial\mathfrak{G}$ and let $f : K \to L$ be a map in dimensions $< q$. If K_n is representable for all $n \geqq q$ and if $H_n(L(M)) = 0$ for all $n \geqq q - 1$ and all $M \varepsilon \mathfrak{M}$, then f admits an extension $f' : K \to L$ (defined in all dimensions). If $f', f'' : K \to L$ are two such extension of f then there is a homotopy $D : f' \simeq f''$ with $D_n = 0$ for all $n < q$.*

3. Groups with operators. Let W be an associative system with a unit element. Each element $w \varepsilon W$ gives rise to a transformation $w : W \to W$ defined by $w(x) = wx$. This gives rise to a category $\mathfrak{A}$ containing one object W and maps $w \varepsilon W$. The object W will be regarded as a model; thus in this case the set $\mathfrak{M}$ contains all the objects of $\mathfrak{A}$.

What is a covariant functor T on $\mathfrak{A}$ with values in $\mathfrak{G}$? It consists of an abelian group $G = T(W)$ and of endomorphisms $T(w) : G \to G$ such that

$T(w_2w_1) = T(w_2)T(w_1)$ and that $T(1) =$ identity. Thus G is an abelian group with W as left operators.

If we now inspect the definition of $\tilde{T}$ we find that $\tilde{G}$ is the free abelian group generated by pairs (w, g), $w \varepsilon W$, $g \varepsilon G$ with operators defined by $w'(w, g) = (w'w, g)$. Thus $\tilde{G}$ is the W-free group with a W-base formed by the elements $g = (1, g)$, $g \varepsilon G$. The map $\Phi: \tilde{G} \to G$ maps the generator g of $\tilde{G}$ into the element g of G. Clearly Φ maps $\tilde{G}$ onto G.

In order that the functor corresponding to G be representable, there must exist a W-map $\Psi: G \to \tilde{G}$ such that $\Phi\Psi =$ identity. Such a Ψ always exists if G is W-free.

These remarks may be more conveniently restated using the algebra Λ of W, i. e. the additive free abelian group generated by the elements $w \varepsilon W$ with a multiplication defined by that of W. Then G becomes a left Λ-module. Natural transformations of functors translate into Λ-homomorphisms. Further, $\tilde{G}$ is the free Λ-module generated by the elements $g \varepsilon G$. The existence of the Λ-homomorphism $\Psi: G \to \tilde{G}$, with $\Phi\Psi =$ identity, is equivalent with the property that G be a projective Λ-module (one of several equivalent definitions: projective = direct summand of a free module).

These remarks and the results of § 2 yield a new proof (in a somewhat more general form) of a basic result concerning complexes with operators [2, § 5].

4. Doubling of subcategories. For the purpose of subsequent applications of the results of § 2 it will be convenient to describe a certain abstract method for constructing categories.

Let $\mathfrak{A}$ be a category and $\mathfrak{B}$ a subcategory of $\mathfrak{A}$. We define a new category $\mathfrak{A}^*$, called the result of *doubling* the subcategory $\mathfrak{B}$, as follows.

Fro each object $B \varepsilon \mathfrak{B}$ we introduce a new object B^*, and for each map $\beta: B \to B'$ in $\mathfrak{B}$ we introduce a new map $\beta^*: B^* \to B'^*$ with $(\beta_2\beta_1)^* = \beta_2{}^*\beta_1{}^*$. These new objects and new maps constitute a category $\mathfrak{B}^*$ isomorphic to $\mathfrak{B}$. The objects in $\mathfrak{A}^*$ are the objects A of $\mathfrak{A}$ and the objects B^* of $\mathfrak{B}^*$. The maps of $\mathfrak{A}^*$ are the maps of $\mathfrak{A}$, the maps of $\mathfrak{B}^*$ (each with the given composition rules), plus new maps $\gamma^\#: B^* \to A$, one for each map $\gamma: B \to A$ in $\mathfrak{A}$ with $B \varepsilon \mathfrak{B}$. The composition rule for these new maps with either previous type are given as follows. If $B' \xrightarrow{\beta} B \xrightarrow{\gamma} A \xrightarrow{\alpha} A'$ with $\beta \varepsilon \mathfrak{B}, \gamma, \alpha \varepsilon \mathfrak{A}$, then $(\gamma\beta)^\# = \gamma^\#\beta^*$, $(\alpha\gamma)^\# = \alpha\gamma^\#$. The axioms for a category are readily verified.

These rules show that for each $\gamma: B \to A$ we have $\gamma^\# = \gamma i_B{}^\#$, where $i_B: B \to B$ is the identity map of B. The map $i_B{}^\#: B^* \to B$ is called the *inclusion* map for B. For $\beta: B' \to B$ we have $\beta^\# = i_B{}^\#\beta^*$. Hence any map

in $\mathcal{A}^*$ is uniquely representable as one of the maps α, β^* or $i_B{}^{\#}$ or their composites, subject to the rule $i_B{}^{\#}\beta^* = \beta i_{B'}{}^{\#}$ for $\beta: B' \to B$ in $\mathcal{B}$.

In the applications, the set of objects B^* will usually be the set $\mathfrak{M}$ of models for the category $\mathcal{A}^*$.

5. Homology theories for multiplicative systems. We shall show here how the results of § 2 can be applied to yield the main results of [3]. We shall use, without explanation, the notation and terminology introduced there.

Let F be a free multiplicative system with generators $g_1, \cdots, g_i, \cdots$ and $\mathfrak{M}(F)$ the category of multiplicative systems belonging to F [3, § 2]. Let Φ be an admissible set of endomorphisms of F. The system $\{F, \Phi\}$ is a subcategory of $\mathfrak{M}(F)$ and we denote by $\mathcal{A}$ the result of doubling up the subcategory $\{F, \Phi\}$ of $\mathfrak{M}(F)$. The new object F^* is chosen as the (only) model.

Let K be a Φ-construction on $\mathfrak{M}(F)$, [3, § 5]. We shall regard K as a covariant functor on $\mathfrak{M}(F)$ with values in $\partial\mathcal{G}$. We shall show how K can be extended to a functor K^* on $\mathcal{A}$. We define $K^*(F^*)$ as the Φ-subcomplex $K(F, \Phi)$ of $K(F)$. The map $K^*(i_F{}^{\#}): K^*(F^*) \to K^*(F)$ is defined as the inclusion map $K(F, \Phi) \to K(F)$. If $\phi \in \Phi$ then $K^*(\phi^*): K^*(F^*) \to K^*(F^*)$ is defined as the endomorphism $K(F, \Phi) \to K(F, \Phi)$ defined by $K(\phi)$. If L is another (augmented) Φ-construction and $f: K \to L$ is an (augmented) map as defined in [3, § 5], then f admits a unique extension $f^*: K^* \to L^*$. Conversely, for each map $f^*: K^* \to L^*$ the restriction $f: K \to L$ is a map in the sense of [3, § 5]. The same applies to homotopies.

We shall now show that each component $K_q{}^*$ of K^* is representable for $q > 0$. (For $q = 0$, $K_0{}^*$ is the augmentation functor, which is not representable). For $M \in \mathfrak{M}(F)$, each q-dimensional cell in $K(M)$ has a form $[x_1, \cdots, x_r]_t$, and is of type t with entries $x_1, \cdots, x_r \in M$. Let $\alpha: F \to M$ be the map defined by $\alpha(g_i) = x_i$ for $i = 1, \cdots, r$ and $\alpha(g_i) = 1$ for $i > r$. Then $[g_1, \cdots, g_r]_t$ is a q-cell of $K^*(F^*)$. The mapping

$$[x_1, \cdots, x_r]_t \to (\alpha^{\#}, [g_1, \cdots, g_r]_t)$$

then defines a representation of K_q. With these preliminaries it is clear that Theorem 6.1 of [3] is a consequence of the results of § 2.

The same remarks apply to the considerations of [2, § 15]. Instead of the single free system F we consider the sequence $\{G_R{}^i\}$ where $G_R{}^i$ is generated by $g_1, \cdots, g_i$ $(i = 0, 1, \cdots)$. In this case, instead of a single model we have a sequence of models.

6. Simplicial singular homology. Let X be a topological space. A singular n-simplex T of X is a function $T(\lambda_0, \cdots, \lambda_n) \varepsilon X$ defined for $0 \leqq \lambda_i$, $\lambda_0 + \cdots + \lambda_n = 1$ and continuous in the topology induced by the cartesian product of the variables. The faces $F_i T$ $(i = 0, \cdots, n)$ are $(n-1)$- simplexes defined as

$$(F_i T)(\lambda_0, \cdots, \lambda_{n-1}) = T(\lambda_0, \cdots, \lambda_{i-1}, 0, \lambda_i, \cdots, \lambda_n).$$

Then

$$F_i F_j = F_{j-1} F_i, \qquad i < j. \tag{6.1}$$

We define $S_n(X)$ as the free group generated by all singular n-simplexes in X. Then

$$\partial T = \sum_{i=0}^{n} (-1)^i F_i T$$

is a homomorphism $\partial : S_n(X) \to S_{n-1}(X)$ with $\partial\partial = 0$. This yields a chain complex $S(X)$. It will be convenient to "augment" $S(X)$ by defining $S_{-1}(X)$ to be the group of integers with $\partial T = 1$ for each 0-simplex. Henceforth $S(X)$ will denote the augmented complex.

If $f : X \to Y$ is a continuous map and T is a singular n-simplex in X, then the composition fT is a singular n-simplex in Y. This yields maps $S_n(f) : S_n(X) \to S_n(Y)$ and $S(f) : S(X) \to S(Y)$ and therefore functors S_n and S defined on the category $\mathfrak{A}$ of topological spaces (with continuous maps) and with values in the categories $\mathfrak{G}$ and $\partial\mathfrak{G}$ respectively.

For each n-simplex T $(n \geqq 0)$ in X and each $i = 0, \cdots, n$ we define the $(n+1)$-simplex $D_i T$ in X as follows:

$$(D_i T)(\lambda_0, \cdots, \lambda_{n+1}) = T(\lambda_0, \cdots, \lambda_{i-1}, \lambda_i + \lambda_{i+1}, \lambda_{i+2}, \cdots, \lambda_n).$$

We have the identities

$$D_i D_j = D_{j+1} D_i, \qquad i \leqq j, \tag{6.2}$$

$$F_i D_j = D_{j-1} F_i,\ i < j;\ F_j D_j = F_{j+1} D_j = I;\ F_i D_j = D_j F_{i-1},\ i > j+1. \tag{6.3}$$

These imply that the simplexes $D_n T$ (where $\dim T = n$) form a subcomplex $\bar{D}S(X)$ of $S(X)$. Similarly the simplexes $D_i T$ $(i = 0, 1, \cdots, \dim T)$ form a subcomplex $DS(X)$. We introduce the quotient functors $\bar{S} = S/\bar{D}S$, $S^N = S/DS$. We call $S(X)$ the *singular complex* of X, $S^N(X)$ the singular complex of X *normalized*, $\bar{S}(X)$ the singular complex of X *normalized at the top*. We observe that in dimensions < 1, the three singular complexes of X coincide.

THEOREM III. *Let $f: S \to S^N$ be the natural factorization homomorphism. There exists then a map $g: S^N \to S$ and homotopies $H: gf \cong$ identity $G: fg \cong$ identity. The same conclusion applies to the natural maps $f_1: S \to \bar{S}$ and $f_2: \bar{S} \to S^N$.*

Proof. In the category $\mathcal{A}$ of topological spaces (on which the functors $S, \bar{S}, S^N$ are defined) we consider the set $\mathfrak{M}$ of *models* consisting of all spaces contractible to a point. The theorem now follows from Theorem II and the following two lemmas:

LEMMA 6.1. *For any model M we have*

$$H_n(S(M)) = H_n(\bar{S}(M)) = H_n(S^N(M)) = 0.$$

LEMMA 6.2. *The functors S_n, $\bar{S}_n$, and $S_n{}^N$ are representable for all n.*

Proof of 6.1. Let I denote the unit interval $0 \leqq \lambda \leqq 1$. Since any model M is contractible, there is a homotopy $H: I \times M \to M$ such that $H(0, x) = x$, $H(1, x) = p$ for all $x \varepsilon M$, where p is a fixed point of M.

For each n-simplex T $(n \geqq 0)$ in M we define the $(n+1)$-simplex hT in M by setting

$$hT(\lambda_0, \cdots, \lambda_{n+1}) = H(\lambda_0, T(\kappa\lambda_1, \cdots, \kappa\lambda_{n+1})), \ \kappa = 1/(1-\lambda_0), \ \lambda_0 \neq 1;$$
$$hT(1, 0, \cdots, 0) = p.$$

For $n = -1$ we define $h(1)$ to be the 0-simplex of M located at p. Since $hD_iT = D_{i+1}hT$ we may regard h as a homomorphism

$$S_n(M) \to S_{n+1}(M), \ \bar{S}_n(M) \to \bar{S}_{n+1}(M), \ S_n{}^N(M) \to S_{n+1}{}^N(M).$$

For $n > 0$ we have $F_0h = I$, $F_ih = hF_{i-1}$ if $i > 0$. For $n = 0$ we have $F_0h = I$, $F_1h = h(1)$. These relations imply $\partial h + h\partial = I$, thus yielding the conclusion of 6.1.

Proof of 6.2. Let Δ^n be the simplex consisting of all points $(\lambda_0, \cdots, \lambda_n)$, $\lambda_i \geqq 0$, $\Sigma\lambda_i = 1$, with the usual topology. Every singular n-simplex T is then a map $T: \Delta^n \to X$. In particular, the identity map $e_n: \Delta^n \to \Delta^n$ is an n-simplex in $S(\Delta^n)$. The correspondence $T \to (T, e_n)$ yields then a representation of the functor S_n for $n \geqq 0$. For $n = -1$ the proof is trivial.

Next we consider the natural factorization maps $\xi_1: S_n \to \bar{S}_n$, $\xi_2: S_n \to S_n{}^N$. The expressions $\eta_1T = (I - D_{n-1}F_n)T$, $\eta_2T = (I - D_0F_1)(I - D_1F_2) \cdots (I - D_{n-1}F_n)T$, satisfy $\eta_1T = 0$ for $T \varepsilon \bar{D}S_{n-1}$ and $\eta_2T = 0$ for $T \varepsilon DS_{n-1}$.

Thus η_1 and η_2 yield maps $\eta_1 : \bar{S}_n \rightarrow S_n$, $\eta_2 : S_n{}^N \rightarrow S_n$. Clearly $\xi_1\eta_1 =$ identity, $\xi_2\eta_2 =$ identity. The representability of $\bar{S}_n$ and $S_n{}^N$ is now a consequence of the following lemma.

LEMMA 6.3. *Let T and T_1 be functors with values in $\mathfrak{G}$ and $\xi : T \rightarrow T_1$, $\eta : T_1 \rightarrow T$ be natural transformations such that $\xi\eta =$ identity. If T is representable then so is T_1.*

Proof. We have the commutative diagram

$$\begin{tikzcd}
\tilde{T}_1 \arrow[r, "\tilde{\eta}"] \arrow[d, "\Phi_1"] & T \arrow[r, "\tilde{\xi}"] \arrow[d, "\Phi"] & \tilde{T}_1 \arrow[d, "\Phi_1"] \\
T_1 \arrow[r, "\eta"'] & T \arrow[r, "\xi"'] & T_1
\end{tikzcd}$$

where $\tilde{\eta}(\phi, m) = (\phi, \eta m)$, $\tilde{\xi}(\phi, m) = (\phi, \xi m)$. Let $\Psi : T \rightarrow \tilde{T}$ be a representation of T and define $\Psi_1 : T_1 \rightarrow \tilde{T}_1$ as $\Psi_1 = \tilde{\xi}\Psi\eta$. Then

$$\Phi_1\Psi_1 = \Phi_1\tilde{\xi}\Psi\eta = \xi\Phi\Psi\eta = \xi\eta = \text{identity},$$

as desired.

Remark. Theorem III is known and has been proved [4] for the more general class of complete semi-simplicial complexes. In this more general case the proof still can be carried out using the method of acyclic models.

7. Cubical singular homology. A singular n-cube R in X is a function $R(\mu_1, \cdots, \mu_n) \in X$ defined for $0 \leqq \mu_i \leqq 1$ and continuous in the topology of the cartesian product of the variables. If $n = 0$, then R is interpreted as a single point of X. The *front* and *aft faces* A_iR and B_iR $(i = 1, \cdots, n)$ are defined as $(n-1)$-cubes

$$(A_iR)(\mu_1, \cdots, \mu_{n-1}) = R(\mu_1, \cdots, \mu_{i-1}, 0, \mu_i, \cdots, \mu_{n-1}),$$

$$(B_iR)(\mu_1, \cdots, \mu_{n-1}) = R(\mu_1, \cdots, \mu_{i-1}, 1, \mu_i, \cdots, \mu_{n-1}).$$

Then

$$\begin{aligned} A_iA_j &= A_{j-1}A_i, \qquad & B_iB_j &= B_{j-1}B_i, \\ A_iB_j &= B_{j-1}A_i, \qquad & B_iA_j &= A_{j-1}B_i. \end{aligned} \qquad i < j \tag{7.1}$$

As before we regard the singular n-cubes in X as generators of a free group $Q_n(X)$ and introduce the operator

$$\partial R = \sum_{i=1}^{n} (-1)^i (A_i R - B_i R).$$

Then (7.1) implies $\partial\partial = 0$. Thus the groups $Q_n(X)$ and the operator ∂ define a chain complex $Q(X)$. As before we augment $Q(X)$ by setting $Q_{-1}(X) =$ integers, $\partial R = 1$ if $\dim R = 0$. Also as before we convert $Q(X)$ into a functor Q defined on $\mathfrak{A}$ with values $\partial\mathfrak{G}$.

For each n-cube R ($n \geqq 0$) in X and each $i = 1, \cdots, n+1$ we define the $(n+1)$-cube $E_i R$ of X as follows

$$(E_i R)(\mu_1, \cdots, \mu_{n+1}) = R(\mu_1, \cdots, \mu_{i-1}, \mu_{i+1}, \cdots, \mu_{n+1}).$$

We have the identities

$$E_i E_j = E_{j+1} E_i, \qquad i \leqq j, \tag{7.2}$$

$$\begin{cases} A_i E_j = E_{j-1} A_i, \quad B_i E_j = E_{j-1} B_i, & i < j; \\ A_j E_j = B_j E_j = I; \\ A_i E_j = E_j A_{i-1}, \quad B_i E_j = E_j B_{i-1}, & i > j. \end{cases} \tag{7.3}$$

These imply that the cubes $E_{n+1} R$ (where $\dim R = n$) form a subcomplex $\bar{E}Q(X)$ of $Q(X)$. Similarly the cubes $E_i R$ ($i = 1, \cdots, 1 + \dim R$) form a subcomplex $EQ(X)$. We introduce the quotient functors $\bar{Q} = Q/\bar{E}Q$, $Q^N = Q/EQ$.

THEOREM IV. *Let $f : \bar{Q} \to Q^N$ be the natural factorization homomorphism. There exist then a map $g : Q^N \to \bar{Q}$ and homotopies $H : gf \simeq$ identity, $G : fg \simeq$ identity.*

The same conclusion does not apply to the maps $Q \to \bar{Q}$ and $Q \to Q^N$.

As in the case of Theorem III, Theorem IV follows from Theorem II and the following two propositions:

PROPOSITION 7.1. *For any model M we have*

$$H_n(\bar{Q}(M)) = H_n(Q^N(M)) = 0.$$

This is not the case for $H_n(Q(M))$.

PROPOSITION 7.2. *The functors Q_n, $\bar{Q}_n$, and $Q_n{}^N$ are representable for all n.*

Proof of 7.1. Using the notation of the proof of 6.1 we define, for each n-cube R of M, an $(n+1)$-cube hR as

$$(hR)(\mu_1,\cdots,\mu_{n+1})=H(1-\mu_1,R(\mu_2,\cdots,\mu_{n+1})), \qquad n\geqq 0$$

For $n=-1$ we define $h(1)$ to be the 0-cube of M located at p. Since $hE_iR=E_{i+1}hR$ we may regard h as an operator $\bar{Q}_n(M)\to\bar{Q}_{n+1}(M)$, $Q_n{}^N(M)\to Q_{n+1}{}^N(M)$. For $n>0$ we have

$$A_1h \in \bar{E}Q_{n-1}(M), \qquad B_1h=I$$

$$A_ih=hA_{i-1}, \qquad B_ih=hB_{i-1} \qquad \text{for } 0<i.$$

For $n=0$ we have $A_1h=h(1)$, $B_1h=I$. These relations imply $\partial h+h\partial\equiv I$ mod $\bar{E}Q(M)$, thus yielding the conclusion of 7.1. The complex $Q(M)$ without any normalization is *not* acyclic, as can be seen in the case in which M consists of a single point.

Proof of 7.2. The proof of the representability of the functor Q_n is the same as that for S_n with the simplex Δ^n replaced by the cube $\square^n$ given by $(\mu_1,\cdots,\mu_n)$, $0\leqq\mu_i\leqq 1$.

The representability of $\bar{Q}_n$ and $Q_n{}^N$ is proved in the same way as for $\bar{S}_n$ and $S_n{}^N$ with the expressions η_1 and η_2 replaced by

$$\eta_1R=(I-E_nA_n)R,\ \eta_2R=(I-E_1A_1)(I-E_2A_2)\cdots(I-E_nA_n)R.$$

Remark. In the simplicial theory the normalizations were essentially a luxury since the functor S already gives the "correct" homology theory. In the cubical theory some normalization is a necessity since the functor Q (without normalization) does not give the "correct" homology groups even in the case of a space consisting of a single point.

8. Comparison of singular and cubical theories. A singular 0-simplex and a singular 0-cube each represent a point of X; thus we are led to identify the functors S_0 and Q_0. Further we identify S_1 with Q_1 by identifying each 1-simplex T with the 1-cube R defined by $R(\mu_1)=T(1-\mu_1,\mu_1)$. These identifications are compatible with the boundary operators $S_1\to S_0$, $Q_1\to Q_0$. These identifications induce identifications $\bar{S}_i=\bar{Q}_i$ and $S_i{}^N=Q_i{}^N$ for $i<2$.

Lemmas 6.1, 6.2, 7.1, 7.2 together with Theorem II yield

THEOREM V. *There exist maps* $f:\bar{S}\to\bar{Q}$, $g:\bar{Q}\to\bar{S}$ *and homotopies* $H:gf\simeq$ *identity*, $G:fg\simeq$ *identity such that* f *and* g *are the identity in*

dimensions < 2 *while* H *and* G *are zero in dimensions* < 2. *The same applies to the pair of functors* S^N, Q^N.

We conclude by giving an explicit form (due to H. Cartan) of a map $f: \bar{S} \to \bar{Q}$ which is the identity in dimensions < 2. We define for each n-simplex T

$$(fT)(\mu_1, \cdots, \mu_n) = T(\lambda_0, \cdots, \lambda_n),$$

where

$$\lambda_0 = 1 - \mu_1, \qquad \lambda_1 = \mu_1(1 - \mu_2), \cdots,$$
$$\lambda_i = \mu_1 \cdots \mu_i(1 - \mu_{i+1}), \quad 0 < i < n, \cdots,$$
$$\lambda_n = \mu_1 \cdots \mu_n.$$

Clearly $\lambda_0 + \cdots + \lambda_i = 1 - \mu_1 \cdots \mu_{i+1}$, $i < n$. Thus $\Sigma\lambda_i = 1$. Further, one easily verifies

$$fF_i = B_{i+1}f, \quad 0 \leqq i < n;$$
$$fF_n = A_n f; \quad A_i f = EA_iA_n f; \quad fD_n = E_{n+1}f.$$

These formulae imply that f maps $\bar{D}S$ into $\bar{E}Q$ and that $\partial f \equiv f\partial \bmod \bar{E}Q$. Thus f induces a map $f: \tilde{S} \to \tilde{Q}$ as desired. It should be further noted that $fT = fT'$ implies $T = T'$, so that f actually yields an isomorphic mapping of $\tilde{S}$ into $\tilde{Q}$.

COLUMBIA UNIVERSITY,
THE UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

[1] Samuel Eilenberg and Saunder MacLane, "General theory of natural equivalences," *Transactions of the American Mathematical Society*, vol. 58 (1945), pp. 231-294.

[2] ———, "Homology of spaces with operators, II," *Transactions of the American Mathematical Society*, vol. 65 (1949), pp. 49-99.

[3] ———, "Homology theories for multiplicative systems," *Transactions of the American Mathematical Society*, vol. 71 (1951), pp. 294-330.

[4] Samuel Eilenberg and J. A. Zilber, "Semi-simplicial complexes and singular homology," *Annals of Mathematics*, vol. 51 (1950), pp. 499-513.

[5] ———, "On products of complexes," *American Journal of Mathematics*, vol. 75 (1953), pp. 200-204.

ON PRODUCTS OF COMPLEXES.*

By SAMUEL EILENBERG and J. A. ZILBER.

The objective of this note is to establish a theorem (stated in § 1) concerning the equivalence, from the point of view of homology, of two kinds of products that may be defined for complete semi-simplicial complexes (see below for a definition). The proof (§ 2) uses the method of acyclic models established in the paper [1] just preceding. Some applications are listed in § 3.

1. The theorem. We write $[m]$ for the set $(0, 1, \cdots, m)$ where m is an integer $\geqq 0$. By a map $\alpha\colon [m] \to [n]$ will be meant a function satisfying $\alpha(i) \leqq \alpha(j)$ for $0 \leqq i \leqq j \leqq m$.

A complete semi-simplicial (abbreviated: c. s. s.) complex K is a collection of "simplexes" σ, to each of which is attached a dimension $q \geqq 0$, such that for each q-simplex σ and each map $\alpha\colon [m] \to [q]$ $(m \geqq 0)$ there is defined an m-simplex $\sigma\alpha$ of K, subject to the conditions

(1) If $\epsilon_q\colon [q] \to [q]$ is the identity then $\sigma\epsilon_q = \sigma$,

(2) If $\beta\colon [n] \to [m]$ then $(\sigma\alpha)\beta = \sigma(\alpha\beta)$.

Let $\epsilon_q{}^i\colon [q-1] \to [q]$ be the map which covers all of $[q]$ except i $(= 0, \cdots, q)$. Then $\sigma\epsilon_q{}^i$ is called the i-th *face* of σ, and the boundary of σ is defined as the chain

$$\partial\sigma = \sum_{i=0}^{q} (-1)^i \sigma\epsilon_q{}^i.$$

If K and L are c. s. s. complexes, a function $f\colon K \to L$ mapping q-simplexes into q-simplexes and such that $f(\sigma\alpha) = (f\sigma)\alpha$ is called a c. s. s. map. For further details see [3, § 8].

Let K and L be two c. s. s. complexes. The *cartesian product* $K \times L$ is a c. s. s. complex whose q-simplexes are pairs (σ, τ) where σ and τ are q-simplexes of K and L respectively. For each map $\alpha\colon [m] \to [q]$ we define $(\sigma, \tau)\alpha = (\sigma\alpha, \tau\alpha)$.

* Received May 26, 1952.

The *tensor* product $K \otimes L$ is an abstract cell complex with r-cells $\sigma \otimes \tau$ where σ is a p-simplex of K, τ a q-simplex of L with $p + q = r$ and

$$\partial(\sigma \otimes \tau) = \partial\sigma \otimes \tau + (-1)^p \sigma \otimes \partial\tau.$$

Both $K \times L$ and $K \otimes L$ may be regarded as chain complexes and may be compared by means of chain transformations and chain homotopies.

THEOREM. *For any two complete semi-simplicial complexes K and L, there exist chain transformations*

$$f : K \times L \to K \otimes L, \qquad g : K \otimes L \to K \times L$$

and chain homotopies

$$D : gf \simeq \textit{identity}, \qquad E : fg \simeq \textit{identity}$$

such that for 0-simplexes $\sigma \in K$, $\tau \in L$,

$$f(\sigma, \tau) = \sigma \otimes \tau, \quad g(\sigma \otimes \tau) = (\sigma, \tau), \quad D(\sigma, \tau) = 0, \quad E(\sigma \otimes \tau) = 0.$$

Moreover, f, g, D and E are natural in the following sense. Let $\phi : K \to K'$, $\psi : L \to L'$ be c. s. s. maps. We consider the induced maps

$$\phi \times \psi : K \times L \to K' \times L', \qquad (\phi \times \psi)(\sigma, \tau) = (\phi\sigma, \psi\tau),$$

$$\phi \otimes \psi : K \otimes L \to K' \otimes L', \qquad (\phi \otimes \psi)(\sigma \otimes \tau) = \phi\sigma \otimes \psi\tau.$$

Then these maps commute properly with f, g, D, E. For example, the diagram

$$\begin{tikzcd} K \times L \arrow[r, "\phi \times \psi"] \arrow[d, "f"] & K' \times L' \arrow[d, "f"] \\ K \otimes L \arrow[r, "\phi \otimes \psi"'] & K' \otimes L' \end{tikzcd}$$

is commutative.

2. Proof of the theorem. For each integer $m \geqq 0$ we define a c. s. s. complex $K[m]$ as follows. A q-simplex of $K[m]$ is any map $\sigma : [q] \to [m]$. For each map $\alpha : [n] \to [q]$, $\sigma\alpha$ is defined as the composite map.

Let $\mathcal{a}$ be the category whose objects are pairs (K, L), where K and L are c. s. s. complexes. A map $(\phi, \psi) : (K, L) \to (K', L')$ *in* $\mathcal{a}$ is a pair of c. s. s. maps $\phi : K \to K'$, $\psi : L \to L'$. Composition is defined by $(\phi', \psi')(\phi, \psi) = (\phi'\phi, \psi'\psi)$ whenever $\phi'\phi$ and $\psi'\psi$ are defined.

In $\mathcal{A}$ we consider the set $\mathfrak{M}$ of models consisting of all pairs $(K[m], K[n])$.

On $\mathcal{A}$ we define two covariant functors P and Q with values in the category of chain complexes as follows. $P(K, L)$ (resp. $Q(K, L)$) is the chain complex obtained from $K \times L$ (resp. $K \otimes L$) by adjoining the group of integers as group of chains in dimension -1 with $\partial(\sigma \times \tau) = 1$ (resp. $\partial(\sigma \otimes \tau) = 1$) for 0-simplexes $\sigma \varepsilon K$, $\tau \varepsilon L$. The maps $P(\phi, \psi)$ (resp. $Q(\phi, \psi)$) are defined as extensions of $\phi \times \psi$ (resp. $\phi \otimes \psi$) obtained by keeping the chains of dimension -1 (i. e. the integers) pointwise fixed.

We first show that for each dimension $r \geqq 0$ the functors P_r and Q_r are representable. If σ is an n-simplex in a c. s. s. complex K, then we denote by ϕ_σ the map $\phi_\sigma : K[n] \to K$ defined for each α in $K[n]$ as $\phi_\sigma \alpha = \sigma\alpha$. In particular $\phi_\sigma \epsilon_n = \sigma$. With these definitions it is clear that the maps

$$\sigma \times \tau \to ((\phi_\sigma, \phi_\tau), \epsilon_r \times \epsilon_r), \ \dim \sigma = \dim \tau = r$$

$$\sigma \otimes \tau \to ((\phi_\sigma, \phi_\tau), \epsilon_p \otimes \epsilon_q), \ \dim \sigma = p, \dim \tau = q, p + q = r$$

yield representations of the functors P_r and Q_r.

Next we prove that the homology groups of the complexes $P(K[m], K[n])$ and $Q(K[m], K[n])$ are all trivial.

For any map $\alpha : [q] \to [r]$ we define a map $F(\alpha) : [q+1] \to [r]$ by setting

$$F(\alpha)(0) = 0, \qquad F(\alpha)(i) = \alpha(i-1) \text{ for } i = 1, \cdots, q+1.$$

Further, we define $\theta_r : [0] \to [r]$ by $\theta_r(0) = 0$.

Then

$$\begin{aligned} &F(\alpha)\epsilon_{q+1}{}^0 = \alpha \\ &F(\alpha)\epsilon_{q+1}{}^i = F(\alpha\epsilon_q{}^{i-1}) && q > 0, i = 1, \cdots, q+1 \\ &F(\alpha)\epsilon_1{}^1 = \theta_r && q = 0. \end{aligned}$$

Next, we define in $P(K[m], K[n])$ and $Q(K[m], K[n])$ homotopy operators G and H as follows:

$$G(\sigma, \tau) = (F(\sigma), F(\tau)), \qquad G(1) = (\theta_m, \theta_n),$$

$$H(\sigma \otimes \tau) = F(\sigma) \otimes \tau \text{ if } \dim \sigma > 0,$$

$$H(\sigma \otimes \tau) = F(\sigma) \otimes \tau + \theta_m \otimes F(\tau) \text{ if } \dim \sigma = 0,$$

$$H(1) = \theta_m \otimes \theta_n.$$

A simple calculation, using the face formulae for $F(\alpha)$ shows that $\partial G + G\partial$

and $\partial H + H\partial$ are identity operators. This proves the assertion concerning the triviality of the homology groups.

The remainder of the proof is now a direct application of Theorem II of [1]. We define the maps $f\colon P \to Q$ and $g\colon Q \to P$ in dimension -1 by $f(1) = 1 = g(1)$ and in dimension zero by

$$f(\sigma, \tau) = \sigma \otimes \tau, \qquad g(\sigma \otimes \tau) = (\sigma, \tau).$$

Then f and g can be extended to maps defined in all dimensions. Since gf and fg coincide with the identity maps in dimensions < 1, the homotopies D and E required by the theorem, also exist in virtue of Theorem II of [1].

Although the proof given here appears to be purely existential, using the representations given for the functors P_r and Q_r and using the homotopies G and H above, explicit formulae for f, g, D and E may be readily found. Such formulae will be found in [2].

3. Applications. Let $X \times Y$ be the cartesian product of two topological spaces X and Y. A q-dimensional singular simplex in $X \times Y$ defines by projection a singular q-simplex in X and one in Y. Conversely a pair of singular q-simplexes one in X and one in Y determine a singular q-simplex in $X \times Y$. It follows that the total singular complex $S(X \times Y)$ (which is a c.s.s. complex; see [3, § 8]), may be identified with the product $S(X) \times S(Y)$. Thus the theorem allows us to assert that from the point of view of homology $S(X \times Y)$ is equivalent with $S(X) \otimes S(Y)$.

Let A and B be subspaces of X and Y respectively. We write $S(X, A)$ for the quotient of $S(X)$ by its subcomplex $S(A)$. Since the maps and homotopies asserted in the theorem are natural, it follows that the relative homology groups

$$H_q(S(X \times Y)/S(A \times Y) \cup S(X \times B)) \tag{1}$$

and

$$H_q(S(X, A) \otimes S(Y, B)) \tag{2}$$

are isomorphic. We consider the triple of complexes

$$(S(X \times Y), S(A \times Y \cup X \times B), S(A \times Y) \cup S(X \times B)).$$

If all the homology groups

$$H_q(S(A \times Y \cup X \times B)/S(A \times Y) \cup S(X \times B)) \tag{3}$$

are trivial, then it follows from the exactness of the homology sequence of the triple above that the groups (1) are isomorphic with

$$H_q(X \times Y, A \times Y \cup X \times B). \tag{4}$$

Thus in this case (2) and (4) are isomorphic.

Our second application concerns the simplicial product of simplicial complexes. Let K and L be simplicial complexes. The *simplicial product* $K \triangle L$ has as vertices pairs (A, B) of vertices $A \varepsilon L$, $B \varepsilon L$. A set $(A^0, B^0), \cdots, (A^n, B^n)$ of vertices of $K \triangle L$ forms a simplex of $K \triangle L$ if and only if $A^0, \cdots, A^n$ are in a simplex of K and $B^0, \cdots, B^n$ are in a simplex of L.

With each simplicial complex K we associate a c. s. s. complex $O(K)$ as follows. The q-simplexes of $O(K)$ are sequences $A^0 \cdots A^q$ of vertices of K contained in a simplex of K. For each map $\alpha \colon [m] \to [q]$ we define $(A^0 \cdots A^q)\alpha = A^{\alpha(0)} \cdots A^{\alpha(m)}$. The homology theories of K and $O(K)$ are equivalent.

With these definitions it is easy to see that $O(K \triangle L) = O(K) \times O(L)$. Thus the theorem of this paper asserts that $O(K \triangle L)$ and $O(K) \otimes O(L)$ are homologically equivalent. It follows that the homology theories of $K \triangle L$ and of $K \otimes L$ (regarding K and L as chain complexes) are equivalent.

This result may be applied in the following situation. Let U and V be coverings of spaces X and Y respectively and let $U \times V$ be the "product" covering of $X \times Y$. Then it is easy to verify the following relation between the nerves of these coverings: $N(U \times V) = N(U) \triangle N(V)$. It follows that that $N(U \times V)$ is homologically equivalent with $N(U) \otimes N(V)$.

COLUMBIA UNIVERSITY,
THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] Samuel Eilenberg and Saunders MacLane, "Acyclic models," *American Journal of Mathematics*, vol. 75 (1953), pp. 189-199.

[2] ———, "On the groups $H(\Pi, n)$, I," to appear in the *Annals of Mathematics*.

[3] Samuel Eilenberg and J. A. Zilber, "Semi-simplicial complexes and singular homology," *Annals of Mathematics*, vol. 51 (1950), pp. 499-513.

ON THE DIMENSION OF PRODUCT SPACES.*

By Kiiti Morita.

Let X and Y be normal spaces of finite dimension. It is well known that the relation

$$\text{(A)} \qquad \dim (X \times Y) \leqq \dim X + \dim Y$$

holds for the three cases: (a) X and Y are separable metric spaces, (b) X and Y are compact (= bicompact) spaces (E. Hemmingsen [6]) and (c) X is an S-space and Y is compact (E. G. Begle [1]). Here a Hausdorff space R is said to be an S-space or to have the star-finite property if every open covering of R has a star-finite refinement (see [1], [3], [9]), and $\dim R \leqq n$ means that every finite open covering of R has a refinement of order not greater than $n+1$.

In the present paper we shall prove the relation (A) for the following three cases:

I. The topological product of X and Y is an S-space.

II. X is a fully normal space and Y is a locally compact fully normal space.

III. X is a countably paracompact normal space and Y is a locally compact metric space.

Separable metric spaces have the Lindelöf property and the Lindelöf property implies the star-finite property for regular spaces [9], and the topological product of an S-space and a compact space is an S-space [1]. Hence the known cases mentioned above are all included in Case I. It is to be noted that a locally compact topological group, regarded as a space, has the star-finite property [9]. Our proof for Case I is based on the fact that if a subspace A of a normal space R is an S-space we have $\dim A \leqq \dim R$.

A topological space R is called paracompact or countably paracompact whenever every open covering or every countable open covering has a locally finite (= neighbourhood-finite in the sense of S. Lefschetz) refinement. It is known that, for Hausdorff spaces, paracompactness is equivalent to full normality [14] and that the topological product of a fully normal space and a

* Received September 29, 1952.

compact normal space is fully normal [2]. Recently it has been proved by C. H. Dowker [5] that (1) a topological space R is a countably paracompact normal space if and only if the topological product $R \times I$ of R with the closed line interval $I = [0, 1]$ is normal, (2) fully normal spaces and perfectly normal spaces are countably paracompact and (3) the topological product of a countably paracompact normal space and a compact metric space is countably paracompact and normal.

It will be shown that for Case II or III the product space $X \times Y$ is a paracompact or countably paracompact normal space respectively. Thus II and III are the important cases for which the topological product $X \times Y$ is normal, and a question raised by E. G. Begle [1] is solved hereby in its wider sense. Our proof of (A) for Case II rests on an addition theorem **(3)** which is a generalization of W. Hurewicz's addition theorem (cf. [10]). As an application of the relation (A) for Case III we note here that the cohomotopy group $\pi^n(X, A)$ in the sense of Borsuk-Spanier [13] can be defined for a countably paracompact normal space X with $\dim (X - A) < 2n - 1$.

The relation stronger than (A) is

$$\text{(B)} \qquad \dim (X \times Y) = \dim X + \dim Y.$$

It will be shown that (B) holds for the following cases:

IV. X is a locally compact fully normal space of dimension $\geqq 0$ and Y is a fully normal space of dimension 1.

V. X is a fully normal space of dimension $\geqq 0$ and Y is a locally finite polytope of dimension $\geqq 0$.

In case X and Y are separable metric spaces the relation (B) for Case IV was established by W. Hurewicz [7]. It follows also from our results that (B) holds for the case where X is a locally compact fully normal space of dimension $\geqq 0$ and Y is an arbitrary (finite or infinite, but non-empty) polytope assigned with the topology due to J. H. C. Whitehead [16].

1. Dimension of subspaces.

THEOREM 1. *If a subspace A of a normal space X is an S-space, then we have* $\dim A \leqq \dim X$.

Proof. Let $\{A \cap G_i \mid i = 1, 2, \cdots, s\}$ be a finite open covering of a subspace A with open sets G_i of X. For each point p of A there exists an open neighbourhood $V(p)$ whose closure is contained in some G_j. Then the family of sets $\{A \cap V(p) \mid p \in A\}$ is an open covering of A and hence, by

assumption, it has a star-finite refinement $\mathfrak{H}$. Because of star-finiteness the open covering $\mathfrak{H}$ can be written in the form $\{A \cap H_{\alpha i} \mid \alpha \varepsilon \Omega, i = 1, 2, \cdots\}$ with open sets $H_{\alpha i}$ of X such that

$$(1) \qquad A \cap H_{\alpha i} \cap H_{\beta j} = 0, \text{ for } \alpha \neq \beta.$$

By construction we have $A \cap H_{\alpha i} \subset A \cap V(p_{\alpha i})$ for some point $p_{\alpha i}$ of A, and hence we may assume without loss of generality that $H_{\alpha i} \subset V(p_{\alpha i})$. It follows then from the property of $V(p)$ that $\bar{H}_{\alpha i} \subset$ some G_j.

Now let $\dim X = n$. Then $\dim \bar{H}_{\alpha i} \leqq n$ and $\bigcup_{i=1}^{\infty} \bar{H}_{\alpha i} \subset \bigcup_{j=1}^{s} G_j$. Hence by the sum theorem (see [6], [10], [15]) there exist open sets $U_{\alpha j}$ $(j = 1, 2, \cdots, s)$ such that

$$(2) \qquad U_{\alpha j} \subset G_j, \qquad j = 1, 2, \cdots, s,$$

$$(3) \qquad \bigcup_{i=1}^{\infty} \bar{H}_{\alpha i} \subset \bigcup_{j=1}^{s} U_{\alpha j},$$

$$(4) \qquad \text{order of } \{U_{\alpha j} \mid j = 1, 2, \cdots, s\} \leqq n + 1.$$

If we put

$$W_j = \bigcup_{\alpha} W_{\alpha j}, \qquad W_{\alpha j} = U_{\alpha j} \cap (\bigcup_{i=1}^{\infty} H_{\alpha i}),$$

then $\{A \cap W_j \mid j = 1, 2, \cdots, s\}$ is an open covering of A and we have

$$(5) \qquad W_j \subset G_j, \qquad j = 1, 2, \cdots, s,$$

$$(6) \qquad \text{order of } \{A \cap W_j \mid j = 1, 2, \cdots, s\} \leqq n + 1,$$

by virtue of (1), (2) and (4). Since $\{A \cap G_j\}$ is an arbitrary open covering of A we see that $\dim A \leqq \dim X$.

Remark. It seems that the relation $\dim A \leqq \dim X$ is not true in general for a fully normal subspace A of a normal space X; for, since a Hausdorff space A of dimension zero in the sense of Menger-Urysohn is imbedded in a compact Hausdorff space X with $\dim X = 0$, the validity of the relation mentioned above implies the equivalence of dimension zero in our sense and dimension zero in the sense of Menger-Urysohn for fully normal spaces, and this equivalence is rather doubtful.

2. The relation (A) for Case I. For the sake of completeness we shall first give a simple proof to the relation (A) for the case (c); the more general Case II will be treated in **5**. Our proof given here, contrary to the existing proofs [1], [6], presupposes no properties of finite polytopes.

Let X be an S-space and Y a compact Hausdorff space, and let $\dim X \leqq m$, $\dim Y \leqq n$. Then any open covering of the product space $X \times Y$ has a refinement $\wp$ of the form

$$(7) \qquad \{G_{\alpha i} \times H(j; \alpha, i) \mid \alpha \in \Omega, i = 1, 2, \cdots; H(j; \alpha, i) \in \mathfrak{H}_{\alpha i}\}$$

where $\{G_{\alpha i} \mid \alpha \in \Omega, i = 1, 2, \cdots\}$ is a star-finite open covering of X of order $\leqq m + 1$ such that $G_{\alpha i} \cap G_{\beta j} = 0$ for $\alpha \neq \beta$, and

$$\mathfrak{H}_{\alpha i} = \{H(j; \alpha, i) \mid j = 1, \cdots, \kappa(\alpha, i)\}$$

is a finite open covering of Y.

For each covering $\mathfrak{H}_{\alpha i}$ we construct a closed covering

$$\{K(j; \alpha, i) \mid j = 1, \cdots, \kappa(\alpha, i)\}$$

of Y such that $K(j; \alpha, i) \subset H(j; \alpha, i)$ and we apply Theorem 3. 4 in our previous paper [10] to a countable number of pairs of sets

$$\{K(j; \alpha, i), H(j; \alpha, i)\}, i = 1, 2, \cdots, j = 1, 2, \cdots, \kappa(\alpha, i),$$

where α is fixed. Then we can find open sets $W(j; \alpha, i)$ of Y such that

$$(8) \qquad K(j; \alpha, i) \subset W(j; \alpha, i) \subset \overline{W(j; \alpha, i)} \subset H(j; \alpha, i),$$

$$(9) \quad \text{order of } \{\overline{W(j; \alpha, i)} - W(j; \alpha, i) \mid i = 1, 2, \cdots; j = 1, \cdots, \kappa(\alpha, i)\} \leqq n.$$

We construct a closed covering $\{F_{\alpha i} \mid \alpha \in \Omega, i = 1, 2, \cdots\}$ of X such that $F_{\alpha i} \subset G_{\alpha i}$ and put

$$\mathfrak{M} = \{F_{\alpha i} \times \overline{V(j; \alpha, i)} \mid \alpha \in \Omega, i = 1, 2, \cdots, j = 1, 2, \cdots, \kappa(\alpha, i)\},$$

where

$$(10) \quad V(j; \alpha, i) = W(j; \alpha, i) - \bigcup_{k=1}^{j-1} \overline{W(k; \alpha, i)}, j \geqq 2; V(1; \alpha, i) = W(1; \alpha, i).$$

Then $\mathfrak{M}$ is a closed covering of $X \times Y$ and

$$F_{\alpha i} \times \overline{V(j; \alpha, i)} \subset G_{\alpha i} \times H(j; \alpha, i).$$

Hence if we prove

$$(11) \qquad \text{order of } \mathfrak{M} \leqq m + n + 1,$$

we have $\dim(X \times Y) \leqq \dim X + \dim Y$ by [11, Theorem 1. 3].

Since $F_{\alpha i} \cap F_{\beta j} = 0$ for $\alpha \neq \beta$, a non-empty intersection of sets of $\mathfrak{M}$ is of the form

$$L = \bigcap_{i=1}^{r} \bigcap_{j=1}^{s_i} (F_{\alpha k_i} \times \overline{V(\tau_i(j); \alpha, k_i)}).$$

where $1 \leqq \tau_i(1) < \cdots < \tau_i(s_i) \leqq \kappa(\alpha, k_i)$, and $1 \leqq k_1 < k_2 < \cdots < k_r$. The order of the covering $\{G_{\alpha i} \mid \alpha \varepsilon \Omega, i = 1, 2, \cdots\}$ being not greater than $m+1$, we have $r \leqq m+1$. On the other hand we have by (10)

$$\bigcap_{i=1}^{r} \bigcap_{j=1}^{s_i} \overline{V(\tau_i(j); \alpha, k_i)} \subset \bigcap_{i=1}^{r} \bigcap_{j=1}^{s_i-1} (\overline{W(\tau_i(j); \alpha, k_i)} - W(\tau_i(j); \alpha, k_i))$$

and hence we obtain $\sum_{i=1}^{r} (s_i - 1) \leqq n$ from (9). Consequently we get

$$\sum_{i=1}^{r} s_i \leqq r + n \leqq m + n + 1,$$

which proves (11).

Thus the relation (A) is established for Case (c).

Now let the topological product of X and Y be an S-space. If we denote Čech's compactifications of X and Y by $\beta(X)$ and $\beta(Y)$ respectively, then we have $\dim X = \dim \beta(X)$, $\dim Y = \dim \beta(Y)$, since by assumption X and Y are S-spaces and hence normal. The relation

$$\dim (\beta(X) \times \beta(Y)) \leqq \dim \beta(X) + \dim \beta(Y)$$

being established above, we obtain from Theorem 1

$$\dim (X \times Y) \leqq \dim (\beta(X) \times \beta(Y)) \leqq \dim X + \dim Y.$$

Hence we have

THEOREM 2. *If X and Y are spaces such that the topological product of X and Y is an S-space, then* $\dim (X \times Y) \leqq \dim X + \dim Y$.

Remark. The space S defined by Sorgenfrey [12] is an S-space with the Lindelöf property such that $S \times S$ is not normal, and moreover $\dim S = 0$. Thus the relation (A) does not hold in general unless $X \times Y$ is normal.

3. An addition theorem. We shall prove an addition theorem which plays an important role in **5**.

THEOREM 3. *Let A_α, $\alpha \varepsilon \Omega$* be closed sets of a normal space X and $\mathfrak{G}$ a locally finite open covering of X. If

1) $(\mathfrak{G})$ — $\dim A_\alpha \leqq n$ for every $\alpha \varepsilon \Omega$,
2) $\dim A_\alpha \cap A_\beta \leqq n-1$ for any distinct $\alpha, \beta \varepsilon \Omega$,
3) there is a locally finite system $\{H_\alpha \mid \alpha \varepsilon \Omega\}$ of open sets of X such that $A_\alpha \subset H_\alpha$ for each α,

then we have $(\mathfrak{G})-\dim \bigcup_{\alpha} A_\alpha \leqq n$. In case X is a fully normal space the condition 3) can be replaced by

3)′ $\{A_\alpha \mid \alpha \varepsilon \Omega\}$ is a locally finite system in X.

Remark. Here by $(\mathfrak{G})-\dim A \leqq n$ we mean that there exists an open covering of a subspace A which is a refinement of $\mathfrak{G}$ and has order $\leqq n+1$.

Proof. The second part of the theorem follows readily from [11, Lemma, p. 22]. We assume that the set Ω of indices α consists of all transfinite ordinals less than a fixed ordinal which will be denoted by α_0, that is, we write $\{A_\alpha \mid \alpha < \alpha_0\}$, and similarly we assume that $\mathfrak{G} = \{G_\lambda \mid \lambda < \lambda_0\}$ where λ_0 is an ordinal. The proof will be carried out along the same line as that of Theorem 2.6 in [10].

Suppose that for any ordinal β less than some ordinal $\alpha < \alpha_0$ we have constructed closed sets $P_{\beta\lambda}(\lambda < \lambda_0)$ such that

$$(\mathrm{T}_\beta) \qquad \begin{aligned} & P_{\beta\lambda} \subset A_\beta \cap G_\lambda, \\ & \bigcup_{\lambda<\lambda_0} \bigcup_{\gamma\leqq\beta} P_{\gamma\lambda} = \bigcup_{\gamma\leqq\beta} A_\gamma, \\ & \text{order of } \{\bigcup_{\gamma\leqq\beta} P_{\gamma\lambda} \mid \lambda < \lambda_0\} \leqq n+1. \end{aligned}$$

By the assumption 1) of the theorem such $P_{\beta\lambda}$ exist certainly for $\beta = 1$. We shall now prove the existence of $P_{\alpha\lambda}$ satisfying (T_α). If we put $Q_\lambda = \bigcup_{\beta<\alpha} P_{\beta\lambda}$, $\lambda < \lambda_0$, then Q_λ are closed, since $P_{\beta\lambda} \subset A_\beta$ and $\{A_\beta\}$ is locally finite by 3). We then have

$$(12) \qquad \text{order of } \{Q_\lambda \mid \lambda < \lambda_0\} \leqq n+1.$$

Because, if there is a point p such that $p \,\varepsilon\, Q_{\lambda_i}$ for $i = 1, 2, \cdots, n+2$, then $p \,\varepsilon\, P_{\beta_i\lambda_i}$ for some $\beta_i < \alpha$ and, since there exists β such that $\beta_i \leqq \beta < \alpha$ for every i, we have $p \,\varepsilon \bigcup_{\gamma\leqq\beta} P_{\gamma\lambda_i}$, for $i = 1, 2, \cdots, n+2$; this contradicts the condition (T_β).

Since $Q_\lambda \subset G_\lambda$, there can be found open sets U_λ, by [11, Theorem 1.3], such that

$$(13) \qquad Q_\lambda \subset U_\lambda \subset G_\lambda,$$

$$(14) \qquad \text{order of } \{U_\lambda \mid \lambda < \lambda_0\} \leqq n+1.$$

On the other hand, it follows from the assumption 1) that there exist open sets V_λ and closed sets $B_\lambda(\lambda < \lambda_0)$ such that

$$(15) \qquad B_\lambda \subset V_\lambda \subset G_\lambda,$$

(16) $$\bigcup_{\lambda<\lambda_0} B_\lambda = A_\alpha,$$

(17) $$\text{order of } \{V_\lambda \mid \lambda < \lambda_0\} \leqq n+1.$$

By the assumptions 2), 3) and the generalized sum theorem [11, Theorem 3. 1] we have

(18) $$\dim A_\alpha \cap (\bigcup_{\beta<\alpha} A_\beta) \leqq n-1.$$

Since $\bigcup_{\beta<\alpha} A_\beta$ are closed, there exists a locally finite system $\{W_\tau \mid \tau < \tau_0\}$ of open sets such that

(19) $A_\alpha \cap (\bigcup_{\beta<\alpha} A_\beta) \subset \bigcup_{\tau<\tau_0} W_\tau,$

(20) $\{\bar{W}_\tau \mid \tau < \tau_0\}$ is a refinement of each of the following coverings:

$$\mathfrak{G}, \{U_\lambda, X-Q_\lambda\}, \{V_\lambda, X-B_\lambda\} \text{ for } \lambda < \lambda_0,$$

(21) order of $\{\bar{W}_\tau \mid \tau < \tau_0\} \leqq n,$

where τ_0 is an ordinal. This is assured by (18), [3, Theorem 3. 5] [11, Theorem 2. 1] and [11, Theorems 1. 2, 1. 3]. Let us put

(22) $$C = \bigcup_{\lambda<\lambda_0} C_\lambda, \qquad C_\lambda = B_\lambda \cap (X - \bigcup_{\tau<\tau_0} W_\tau)$$

and

$$M = \bigcup_{\beta<\alpha} A_\beta = \bigcup_{\lambda} Q_\lambda.$$

Then C and C_λ are closed sets and $M \cap C = 0$. Since $\{X-M, X-C\}$ can be regarded as an open covering of a normal space $\bar{W}_1$ there are closed sets E_1 and F_1 such that

(23) $$\bar{W}_1 = E_1 \cup F_1, \qquad E_1 \subset X - C, \qquad F_1 \subset X - M.$$

Then by an inductive process we can construct closed sets E_τ, F_τ $(\tau < \tau_0)$ so that

(24) $$\bar{W}_\tau = E_\tau \cup F_\tau, \qquad E_\tau \subset X - \bigcup_{\rho<\tau} (E_\rho \cap F_\rho) \cup C, \qquad F_\tau \subset X - M.$$

Because, in case we have constructed E_τ, F_τ satisfying (24) for every τ less than some ordinal $\sigma < \tau_0$, $\bigcup_{\tau<\sigma} (E_\tau \cap F_\tau)$ is closed since $\{W_\tau\}$ is locally finite, and $(\bigcup_{\tau<\sigma} (E_\tau \cap F_\tau) \cup C) \cap M = 0$, and hence the existence of E_σ, F_σ satisfying the condition analogous to (24) can be verified.

By the above construction we have

(25) $$M \cap C = 0, \qquad E_\tau \cap C = 0, \qquad F_\tau \cap M = 0 \text{ for } \tau < \tau_0,$$

$$(26) \qquad (E_\rho \cap F_\rho) \cap (E_\tau \cap F_\tau) = 0 \text{ for } \rho \neq \tau.$$

We prove

$$(27) \qquad \text{order of } \{E_\tau, F_\tau \mid \tau < \tau_0\} \leqq n + 1.$$

Let

$$L = \bigcap_{i=1}^{r} E_{\rho_i} \cap \left(\bigcap_{j=1}^{s} F_{\tau_j}\right) \neq 0$$

and let us denote by $\sigma_1, \cdots, \sigma_t$ the distinct ordinals among $\rho_1, \cdots, \rho_r$, $\tau_1, \cdots, \tau_s$. Then we have $L \subset \bigcap_{i=1}^{t} \bar{W}_{\sigma_i}$ and hence $t \leqq n$ by (21). If there were two distinct σ_i, σ_j which are contained in $\{\rho_1, \cdots, \rho_r\}$ as well as in $\{\tau_1, \cdots, \tau_s\}$, then we would have $L \subset E_{\sigma_i} \cap F_{\sigma_i} \cap E_{\sigma_j} \cap F_{\sigma_j}$, contrary to (26). Thus we have $r + s - 1 \leqq t$ and hence $r + s \leqq t + 1 \leqq n + 1$, which proves (27).

Now let us denote by E_λ' the sum of the sets E_τ which do not intersect Q_μ for every μ less than λ and do intersect Q_λ, and by F_λ' the sum of the sets F_τ which do not intersect C_μ for any $\mu < \lambda$ and do intersect C_λ. The family of sets E_τ (or F_τ) which do not intersect M (or C) shall be denoted by $\{E_\rho'' \mid \rho < \rho_0\}$ (or $\{F_\sigma'' \mid \sigma < \sigma_0\}$), where ρ_0, σ_0 denote some ordinals. Then

$$(28) \qquad E_\rho'' \cap (M \cup C) = 0, \qquad F_\sigma'' \cap (M \cup C) = 0$$

and by (20) we have

$$(29) \qquad Q_\lambda \cup E_\lambda' \subset U_\lambda, \qquad C_\lambda \cup F_\lambda' \subset V_\lambda.$$

Let

$$\mathfrak{M} = \{Q_\lambda \cup (A_\alpha \cap E_\lambda'), C_\lambda \cup (A_\alpha \cap F_\lambda'),$$
$$A_\alpha \cap E_\rho'', A_\alpha \cap F_\sigma'' \mid \lambda < \lambda_0, \rho < \rho_0, \sigma < \sigma_0\}.$$

Then $\mathfrak{M}$ is a closed covering of $M \cup A_\alpha = \bigcup_{\gamma \leqq \alpha} A_\gamma$, since

$$\begin{aligned}
&\bigcup_\lambda (Q_\lambda \cup (A_\alpha \cap E_\lambda')) \cup (\bigcup_\lambda (C_\lambda \cup (A_\alpha \cap F_\lambda'))) \\
&\qquad \cup (A_\alpha \cap (\bigcup_\rho E_\rho'')) \cup (A_\alpha \cap (\bigcup_\sigma F_\sigma'')) \\
&= \bigcup_\lambda Q_\lambda \cup (\bigcup_\lambda C_\lambda) \cup (A_\alpha \cap (\bigcup_\tau E_\tau) \cup (A_\alpha \cap (\bigcup_\tau F_\tau)) \\
&= M \cup C \cup (A_\alpha \cap (\bigcup_\tau \bar{W}_\tau)) = M \cup (A_\alpha \cap (X - \cup W_\tau)) \\
&\qquad \cup (A_\alpha \cap (\cup \bar{W}_\tau)) = M \cup A_\alpha.
\end{aligned}$$

To prove

(30) $$\text{order of } \mathfrak{M} \leqq n+1,$$

suppose that

$$L = \bigcap_{i=1}^{r} [Q_{\lambda_i} \cup (A_\alpha \cap E_{\lambda_i}')] \cap [\bigcap_{j=1}^{s} (C_{\mu_j} \cup (A_\alpha \cap F_{\mu_j}'))]$$
$$\cap (\bigcap_{i=1}^{u} E_{\rho_i}'') \cap (\bigcap_{j=1}^{v} F_{\sigma_j}'') \cap A_\alpha \neq 0.$$

Then it is sufficient to prove $r+s+u+v \leqq n+1$. We distinguish three cases.

Case 1). $u \geqq 1$. By (28) we have

$$L = \bigcap_{i,j} (E_{\lambda_i}' \cap F_{\mu_j}') \cap [\bigcap_{i,j} (E_{\rho_i}'' \cap F_{\sigma_j}'')] \cap A_\alpha$$

and hence $r+s+u+v \leqq n+1$ by (27).

Case 2). $v \geqq 1$. Similarly as in Case 1).

Case 3). $u = v = 0$. In this case we have

$$L = \bigcap_{i=1}^{r} (Q_{\lambda_i} \cup (A_\alpha \cap E_{\lambda_i}')) \cap [\bigcap_{j=1}^{s} (C_{\mu_j} \cup (A_\alpha \cap F_{\mu_j}'))].$$

Case $3)_1$. $r \geqq 1$, $s \geqq 1$. Then by (25) we have

$$L = A_\alpha \cap [\bigcap_{i,j} (E_{\lambda_i}' \cap F_{\mu_j}')] \neq 0$$

and hence $r+s \leqq n+1$ by (27).

Case $3)_2$. $r = 0$. We have then by (29) $L \subset \bigcap_{j=1}^{s} V_{\mu_j}$ and hence $s \leqq n+1$ by (17).

Case $3)_3$. $s = 0$. Similarly as in Case $3)_2$.

Thus (30) is proved.

We put

$$P_{\alpha\lambda} = (A_\alpha \cap E_\lambda') \cup ((\textstyle\bigcup' E_\rho'') \cap A_\alpha) \cup ((\textstyle\bigcup'' F_\sigma'') \cap A_\alpha),$$

where $\bigcup'$ means the sum extending over all ρ such that λ is the least ordinal for which $E_\rho'' \subset G_\lambda$, and the meaning of $\bigcup''$ is analogous. Then $P_{\alpha\lambda}$ $(\lambda < \lambda_0)$ are closed and satisfy the condition (T_α), and the existence of $P_{\alpha\lambda}$ satisfying (T_α) is thus established for any α by transfinite induction.

Let us put finally

$$P_\lambda = \bigcup_{\alpha < \alpha_0} P_{\alpha\lambda}.$$

Then we can prove

$$\text{order of } \{P_\lambda \mid \lambda < \lambda_0\} \leqq n+1$$

similarly as in the proof of (12). Since $P_\lambda \subset G_\lambda$ and $\{P_\lambda\}$ is a closed covering of $\bigcup_\alpha A_\alpha$, we obtain $(\mathfrak{S})$ — $\dim \bigcup_{\alpha<\alpha_0} A_\alpha \leqq n$. This completes our proof.

4. Infinite polytopes. As a preparation to the following section we shall discuss some properties of infinite polytopes.

Let K be an abstract simplicial complex, finite or infinite; local finiteness is not assumed here. Let $\bar{K}$ be the polytope corresponding to this complex; the space $\bar{K}$ is a collection of closed Euclidean simplexes, each corresponding to a simplex of K (or an affine realization [8, p. 6]), and the topology of $\bar{K}$ is defined by the method of J. H. C. Whitehead [16]:

1) Each finite subpolytope has the usual Euclidean topology.

2) A subset of $\bar{K}$ is defined to be open if its intersection with every finite subpolytope $\bar{L}$ is open in the ordinary Euclidean topology of $\bar{L}$.

This topology is finer than (in general, but equivalent to in the case of locally finite complexes) the topology induced by the natural metric as well as the topology of geometric complex (cf. [8]).[1]

According to [16, Theorem 35] $\bar{K}$ is a normal space and every open covering of $\bar{K}$ has a refinement which consists of open stars $O(a, K^*)$ of all vertices of a suitable simplicial subdivision K^* of K. By [8, p. 37] the open covering $\{O(a, K^*)\}$ is point-finite and analytic. According to [4, Theorem 1 and Corollary 3], a point-finite open covering of a normal space is analytic if and only if it has a locally finite refinement (and hence it is a normal covering in the sense of J. W. Tukey). Therefore $\bar{K}$ is paracompact or equivalently fully normal. If the (combinatorial) dimension m of K, the least upper bound of dimensions of all simplexes in K, is finite, then the above consideration shows at the same time that $\dim \bar{K} \leqq \dim K$, and, since K contains an m-simplex, we have $\dim \bar{K} \geqq \dim K$ and hence $\dim \bar{K} = \dim K$. Thus

LEMMA 1. *Any polytope $\bar{K}$ is a fully normal space and the topological dimension of $\bar{K}$ coincides with the combinatorial dimension of K.*

---

[1] Cf. also C. H. Dowker, "Topology of metric complexes," *American Journal of Mathematics*, vol. 74 (1952), pp. 555-577, which arrived at our university after the present work was completed.

Now let $\dim K = \dim \bar{K} = m$ be finite and let $\bar{K}'$ be a barycentric subdivision of $\bar{K}$. Here we consider $\bar{K}$ to be a polytope determined by an affine realization of K and make use of barycentric coordinates (see [8, pp. 6-8]). If we denote by $N(a, K')$ the closed star of a, that is, the sum of closed simplexes of K' which contain a, and by $\{a_\alpha \mid \alpha \varepsilon \Omega\}$ the totality of all vertices of K, then the family $\{N(a_\alpha, K')\}$ is a closed covering of $\bar{K}$. We prove

LEMMA 2. *$\{N(a_\alpha, K')\}$ is a locally finite closed covering of K in case the dimension of K is finite.*

To prove Lemma 2, let x be any point of $\bar{K}$ with the barycentric coordinates $\{x_\alpha \mid \alpha \varepsilon \Omega\}$. Then the set

$$V(x) = \{y \mid \sum_\alpha (x_\alpha - y_\alpha)^2 < 1/(m+1)^2\}$$

is easily shown to be an open neighbourhood of x. If x is contained in an open simplex $\sigma^r = (a_{\alpha_0}, a_{\alpha_1}, \cdots, a_{\alpha_r})$ of K, we have $V(x) \cap N(a_\beta, K') = 0$ for any vertex a_β which does not belong to the simplex σ^r. Because, for such β every point y of $V(x)$ has the coordinate $y_\beta < 1/(m+1)$, while for any point z of $N(a_\beta, K')$ we have $z_\beta \geqq 1/(m+1)$ as is shown by a simple calculation. This proves Lemma 2.

Let X be a fully normal space and $\mathfrak{U} = \{U_\alpha \mid \alpha \varepsilon \Omega\}$ a locally finite open covering of X. Then there exists an open covering $\{V_\alpha \mid \alpha \varepsilon \Omega\}$ such that $\bar{V}_\alpha \subset U_\alpha$ for every α. For each α there exists a non-negative bounded continuous function $f_\alpha(x)$ of X such that $f_\alpha(x) = 1$ for $x \varepsilon \bar{V}_\alpha$ and $f_\alpha(x) = 0$ for $x \varepsilon X - U_\alpha$. Let $N(\mathfrak{U})$ be the nerve of $\mathfrak{U}$ and let us denote by u_α the vertex of $N(\mathfrak{U})$ corresponding to U_α of $\mathfrak{U}$. Then a mapping ϕ of X into the polytope $\overline{N(\mathfrak{U})}$ defined by

$$\phi\colon x \to y = \{y_\alpha\}, \qquad y_\alpha = f_\alpha(x) / \sum_\beta f_\beta(x),$$

where $\{y_\alpha\}$ means the barycentric coordinates of a point y of $\overline{N(\mathfrak{U})}$, is continuous. Because for a point x_0 of X there exists an open neighbourhood $V(x_0)$ such that $V(x_0)$ meets only a finite number of sets of $\mathfrak{U}$; these sets will be denoted by U_{α_i}, $i = 0, 1, \cdots, r$. Then ϕ maps $V(x_0)$ into a closed simplex $\bar{\sigma}^r$ of $N(\mathfrak{U})$ whose vertices are u_{α_i}, $i = 0, 1, \cdots, r$. Since a partial mapping $\phi \mid V(x_0) : V(x_0) \to \bar{\sigma}^r$ is evidently continuous, ϕ itself is continuous. It is easy to see that $\phi^{-1}(O(u_\alpha, N(\mathfrak{U}))) \subset U_\alpha$ for each α, that is, ϕ is a canonical mapping with respect to $\mathfrak{U}$.

5. The relation (A) for Case II.

THEOREM 4. *Let X be a fully normal space and Y a locally compact fully normal space. Then the topological product of X and Y is fully normal, and*

$$\text{(A)} \qquad \dim (X \times Y) \leqq \dim X + \dim Y.$$

Proof. 1) First we shall prove (A) for the case where Y is compact. In case $\dim X = 0$, X is an S-space and (A) holds for this case by Theorem 2; but in this case the covering $\wp$ defined below (see (31)) has order $\leqq n+1$ and thus we have a direct proof. Suppose that (A) has been proved for at most $(m-1)$-dimensional X. Under this assumption of induction we shall prove (A) for an m-dimensional X. We let $\dim Y = n$.

Any open covering $\mathfrak{G}$ of the product space $X \times Y$ has, as is well known, a refinement $\wp$ of the form

$$\text{(31)} \qquad \{U_\alpha \times W_{\alpha_\iota} \mid \alpha \varepsilon \Omega, W_{\alpha_\iota} \varepsilon \mathfrak{W}_\alpha\}$$

where $\mathfrak{U} = \{U_\alpha \mid \alpha \varepsilon \Omega\}$ is a locally finite open covering of X of order $\leqq m+1$ and $\mathfrak{W}_\alpha$ is a finite open covering of Y of order $\leqq n+1$ for each α (cf. [3, p. 220]). Let $N(\mathfrak{U})$ be the nerve of $\mathfrak{U}$ and let us make use of the same notations concerning $N(\mathfrak{U})$ as in the preceding section **4**. We put further $K = N(\mathfrak{U})$. As is shown in **4** there is a continuous mapping ϕ of X into $\bar{K}$ such that $\phi^{-1}(O(u_\alpha, K)) \subset U_\alpha$ for every α. Then a mapping Φ of $X \times Y$ into $\bar{K} \times Y$ defined by $\Phi(x, y) = (\phi(x), y)$, for $x \varepsilon X$, $y \varepsilon Y$ is clearly a continuous mapping and

$$\text{(32)} \qquad \Phi^{-1}(O(u_\alpha, K) \times W_{\alpha_\iota}) \subset U_\alpha \times W_{\alpha_\iota}.$$

If we denote by $\bar{K}'$ the barycentric subdivision of $\bar{K}$, then by Lemma 2 in **4** $\{N(u_\alpha, K') \mid \alpha \varepsilon K\}$ is a locally finite closed covering of $\bar{K}$. Hence, since $N(u_\alpha, K') \subset O(u_\alpha, K)$, there is a locally finite open covering $\{H_\alpha \mid \alpha \varepsilon \Omega\}$ of $\bar{K}$ such that $N(u_\alpha, K') \subset H_\alpha \subset O(u_\alpha, K)$, by virtue of Lemma 1 and [11, Lemma, p. 22]. Then the family

$$\mathfrak{M} = \{H_\alpha \times W_{\alpha_\iota} \mid \alpha \varepsilon \Omega, W_{\alpha_\iota} \varepsilon \mathfrak{W}_\alpha\}$$

is a locally finite open covering of $\bar{K} \times Y$. Since $N(u_\alpha, K') \subset H_\alpha$ and $\mathfrak{W}_\alpha$ has order $\leqq n+1$, we have

$$\text{(33)} \qquad (\mathfrak{M}) - \dim [N(u_\alpha, K') \times Y] \leqq n.$$

On the other hand, in case $\alpha \neq \beta$ the intersection $N(u_\alpha, K') \cap N(u_\beta, K')$

is a sum of simplexes of dimension $\leqq m-1$. Hence by Lemma 1 we have $\dim N(u_\alpha, K') \cap N(u_\beta, K') \leqq m-1$ and consequently by the assumption of induction we get

$$(34) \qquad \dim [(N(u_\alpha, K') \times Y) \cap (N(u_\beta, K') \times Y)] \leqq m+n-1.$$

Therefore, since $\bar{K} \times Y$ is fully normal, we obtain from (33), (34) and Theorem 3

$$(\mathfrak{M}) - \dim \bar{K} \times Y \leqq m+n.$$

Thus there exists an open covering $\{V_\gamma\}$ of $\bar{K} \times Y$ which is a refinement of $\mathfrak{M}$ and has order $\leqq m+n+1$. Then $\{\Phi^{-1}(V_\gamma)\}$ is clearly a refinement of $\wp$ in view of (32) and has order $\leqq m+n+1$. Hence we have $(\mathfrak{G}) - \dim (X \times Y) \leqq m+n$. Consequently the relation (A) is established by induction for the case where Y is compact.

2) Let Y be a locally compact fully normal space. Then there exists a star-finite open covering $\{G_\gamma \mid \gamma \varepsilon \Gamma\}$ of Y such that the closure of each G_γ is compact. Let $\mathfrak{U}$ be any open covering of $X \times Y$. Then for each γ there is a locally finite open covering $\mathfrak{U}_\gamma$ of $X \times \bar{G}_\gamma$ which is a refinement of $\mathfrak{U}$. If we denote by $\mathfrak{V}_\gamma$ the family $\{(X \times G_\gamma) \cap U \mid U \varepsilon \mathfrak{U}_\gamma\}$ it is easy to see that the collection $\mathfrak{V}$ of all $\mathfrak{V}_\gamma$: $\mathfrak{V} = \{V \mid V \varepsilon \mathfrak{V}_\gamma, \gamma \varepsilon \Gamma\}$ is a locally finite open covering of $X \times Y$. Since $\mathfrak{V}$ is clearly a refinement of $\mathfrak{U}$, this proves that the product space $X \times Y$ is fully normal.

The covering $\{X \times \bar{G}_\gamma \mid \gamma \varepsilon \Gamma\}$ is locally finite and by the proof given in 1) we have

$$\dim (X \times \bar{G}_\gamma) \leqq \dim X + \dim \bar{G}_\gamma \leqq \dim X + \dim Y.$$

Hence from the generalized sum theorem [11, Theorem 3. 2] it follows that $\dim (X \times Y) \leqq \dim X + \dim Y$.

Thus the theorem is completely proved.

6. The relation (A) for Case III.

THEOREM 5. *If X is a countably paracompact normal space and Y a locally compact metric space, then the topological product $X \times Y$ of X with Y is a countably paracompact normal space and the relation* $\dim (X \times Y) \leqq \dim X + \dim Y$ *holds.*

Proof. 1) First we shall deal with the case where Y is compact. Let $\dim X = m$, $\dim Y = n$ and $s = m+n+1$. Let $P^{(i)}, Q^{(i)}$ $(i = 1, 2, \cdots, s)$ be s pairs of closed sets of the product space $X \times Y$ such that $P^{(i)} \cap Q^{(i)} = 0$

for $i=1,2,\cdots,s$. We shall prove that there exist open sets $U^{(i)}$, $i=1,\cdots,s$ of $X\times Y$ such that

$$P^{(i)}\subset U^{(i)}\subset \bar{U}^{(i)}\subset X\times Y-Q^{(i)},\qquad i=1,2,\cdots,s, \tag{35}$$

$$\bigcap_{i=1}^{s} B(U^{(i)})=0. \tag{36}$$

Here $B(A)$ means the boundary of a set A.

Since Y is a compact metric space, there is a countable basis for open sets of Y. We take s arbitrary countable bases $\mathfrak{L}^{(i)}$, $i=1,2,\cdots,s$ of Y and denote by $\{M_j^{(i)},L_j^{(i)}\}$, $j=1,2,\cdots$ the totality of pairs of sets of $\mathfrak{L}^{(i)}$ such that $\bar{M}_j^{(i)}\subset L_j^{(i)}$. We now apply our Theorem 3.4 in [10] to

$$\{\bar{M}_j^{(i)},L_j^{(i)}\},\qquad i=1,2,\cdots,s,\ j=1,2,\cdots.$$

Then we can find open sets $G_j^{(i)}$ such that $\bar{M}_j^{(i)}\subset G_j^{(i)}\subset L_j^{(i)}$ and

(37) order of the family $\{B(G_j^{(i)})\mid i=1,2,\cdots,s;j=1,2,\cdots\}\leqq n$.

We note that $\{G_j^{(i)}\mid j=1,2,\cdots\}$ is a countable basis of Y for each i.

Denoting by Γ the totality of all finite subsets of the set of natural numbers, we put $H_\gamma^{(i)}=\bigcup_{j\varepsilon\gamma}G_j^{(i)}$, for $\gamma\varepsilon\Gamma$. For each point x of X let $P_x^{(i)}$ be the closed set of Y defined by $x\times P_x^{(i)}=(x\times Y)\cap P^{(i)}$; similarly let $x\times Q_x^{(i)}=(x\times Y)\cap Q^{(i)}$. Let us put further

$$V_\gamma^{(i)}=\{x\mid P_x^{(i)}\subset H_\gamma^{(i)}\subset \bar{H}_\gamma^{(i)}\subset Y-Q_x^{(i)}\}. \tag{38}$$

Then $\{V_\gamma^{(i)}\mid\gamma\varepsilon\Gamma\}$ is an open covering of X as is shown by C. H. Dowker [5, p. 222]. Since by assumption X is countably paracompact and the cardinal number of Γ is countable, there is a locally finite countable open covering $\{W_\gamma^{(i)}\mid\gamma\varepsilon\Gamma\}$ of X such that $W_\gamma^{(i)}\subset V_\gamma^{(i)}$ for $\gamma\varepsilon\Gamma$. It follows further that there is a closed covering $\{F_\gamma^{(i)}\mid\gamma\varepsilon\Gamma\}$ such that $F_\gamma^{(i)}\subset W_\gamma^{(i)}$. We now apply our Theorem 3.4 in [10] again to

$$\{F_\gamma^{(i)},W_\gamma^{(i)}\},\qquad i=1,2,\cdots,s;\gamma\varepsilon\Gamma.$$

Then we can find open sets $U_\gamma^{(i)}$ of X such that

$$F_\gamma^{(i)}\subset U_\gamma^{(i)}\subset W_\gamma^{(i)}, \tag{39}$$

(40) order of $\{B(U_\gamma^{(i)})\mid i=1,2,\cdots,s;\gamma\varepsilon\Gamma\}\leqq m$.

Now let us put $U^{(i)}=\bigcup_{\gamma\varepsilon\Gamma}(U_\gamma^{(i)}\times H_\gamma^{(i)})$. Since $\{U_\gamma^{(i)}\times H_\gamma^{(i)}\mid\gamma\varepsilon\Gamma\}$ is locally finite, we have $\bar{U}^{(i)}=\bigcup_\gamma\overline{U_\gamma^{(i)}\times H_\gamma^{(i)}}$ and hence

(41) $$B(U^{(i)}) \subset \bigcup_{\gamma \varepsilon \Gamma} B(U_\gamma{}^{(i)} \times H_\gamma{}^{(i)}).$$

For any point (x, y) of $P^{(i)}$ we have $x \varepsilon U_\gamma{}^{(i)}$ for some $\gamma \varepsilon \Gamma$ and hence $(x, y) \varepsilon U_\gamma{}^{(i)} \times H_\gamma{}^{(i)}$, and consequently

(42) $$P^{(i)} \subset U^{(i)}.$$

On the other hand, $\overline{U_\gamma{}^{(i)} \times H_\gamma{}^{(i)}} \cap Q^{(i)} = (\bar{U}_\gamma{}^{(i)} \times \bar{H}_\gamma{}^{(i)}) \cap Q^{(i)} = 0$, and hence

(43) $$\bar{U}^{(i)} \subset X \times Y - Q^{(i)}.$$

We now prove

(44) $$\bigcap_{i=1}^{s} B(U^{(i)}) = 0.$$

For this purpose it is sufficient, in view of (41), to prove that

$$L = \bigcap_{i=1}^{s} B(U_{\gamma_i}{}^{(i)} \times H_{\gamma_i}{}^{(i)}) = 0, \text{ for } \gamma_i \varepsilon \Gamma.$$

Denoting by Δ the family of all subsets δ of $\{1, 2, \cdots, s\}$ we have

$$L = \bigcup_{\delta \varepsilon \Delta} (E_\delta \cap F_\delta),$$

where

$$E_\delta = \bigcap_{i \varepsilon \delta} (B(U_{\gamma_i}{}^{(i)}) \times \bar{H}_{\gamma_i}{}^{(i)}), \qquad F_\delta = \bigcap_{i \varepsilon \bar{\delta}} (\bar{U}_{\gamma_i}{}^{(i)} \times B(H_{\gamma_i}{}^{(i)})).$$

In case the cardinal number of δ is greater than or equal to $m + 1$, we see by (40) that $E_\delta = 0$, and in case the cardinal number of δ is less than $m + 1$, we have $F_\delta = 0$ by (37) since $B(H_\gamma{}^{(i)}) \subset \bigcup_{j \varepsilon \gamma} B(G_j{}^{(i)})$. Hence we have $E_\delta \cap F_\delta = 0$ in every case and consequently $L = 0$. This proves (44).

Thus the existence of open sets $U^{(i)}$ satisfying (35), (36) is established, and hence, according to a generalization of Eilenberg-Otto's theorem (see [6], [10]) we see that $\dim (X \times Y) \leqq \dim X + \dim Y$ for a compact metric space Y.

2) Let Y be a locally compact metric space. Then there exists a star-finite open covering $\{G_\alpha \mid \alpha \varepsilon \Omega\}$ of Y such that the closure of each G_α is compact [9]. Similarly as in 2) of the proof of Theorem 4 we can prove that the product space $X \times Y$ is countably paracompact.

We next prove the normality of $X \times Y$. Let P, Q be disjoint closed sets of $X \times Y$. Since $X \times \bar{G}_\alpha$ is normal [5], there exists an open set H_α of $X \times \bar{G}_\alpha$ such that $P \cap (X \times \bar{G}_\alpha) \subset H_\alpha$, $\bar{H}_\alpha \cap Q \cap (X \times \bar{G}_\alpha) = 0$. If we put $K_\alpha = H_\alpha \cap (X \times G_\alpha)$, then K_α is an open set of $X \times Y$ and we have

$P \cap (X \times G_\alpha) \subset K_\alpha$, $\bar{K}_\alpha \cap Q = 0$, and hence $P \subset K$, $\bar{K} \cap Q = 0$ where $K = \bigcup_\alpha K_\alpha$, since $\{K_\alpha\}$ is locally finite. This proves that $X \times Y$ is normal.

Finally we construct a closed covering $\{F_\alpha \mid \alpha \varepsilon \Omega\}$ of Y such that $F_\alpha \subset G_\alpha$. Then, as is proved in 1), we have

$$\dim (X \times F_\alpha) \leqq \dim X + \dim F_\alpha \leqq \dim X + \dim Y.$$

Since $X \times F_\alpha \subset X \times G_\alpha$ and $\{X \times G_\alpha\}$ is locally finite, we have $\dim (X \times Y) \leqq \dim X + \dim Y$ by virtue of the generalized sum theorem [11, Theorem 3.1]. Thus Theorem 5 is completely proved.

7. The relation (B) for Cases IV and V. Let X be a fully normal space and let $\dim X = n$ (n finite). In the following we assume $n \geqq 1$ and apply Hurewicz's method [7]; in case $n = 0$ the relation (B) is clearly true. Since $\dim X = n$, there exist a closed set A and a continuous mapping f of A into an $(n-1)$-sphere S^{n-1} such that f is not extensible over X (see [3], [6], [10]). Assuming that S^{n-1} is a boundary of an n-simplex σ^n we extend f to a continuous mapping F from X into $\bar{\sigma}^n$. There is no loss of generality in assuming that $A = F^{-1}(S^{n-1})$. We denote by X_0 the space obtained from X by contracting the closed set A to a point p_0 (that is, X_0 is the decomposition space determined by the decomposition $X = \bigcup_{x \varepsilon X-A} x \cup A$) and by S_0^n the space obtained from $\bar{\sigma}^n$ by contracting S^{n-1} to a point q_0; the continuous mappings associated with these decompositions will be denoted by $\phi : X \to X_0$; $\psi : \bar{\sigma}^n \to S_0^n$ respectively. S_0^n is clearly homeomorphic to an n-sphere and X_0 is fully normal. If we define a continuous mapping f_0 of X_0 into S_0^n by $f_0(x) = \psi(F(\phi^{-1}(x)))$ for $x \varepsilon X_0 - p_0$, $f_0(p_0) = q_0$, then f_0 is an essential mapping of X_0 into S_0^n. This result is established by C. H. Dowker [3, p. 235].

We observe that $\dim X_0 = n$. Since $A = F^{-1}(S^{n-1})$, A is a G_δ-set of X and hence there exist a countable number of closed sets A_i, $i = 1, 2, \cdots$ such that

$$X - A = \bigcup_{i=1}^{\infty} A_i. \tag{45}$$

Since $\dim A_i \leqq n$ for every i and

$$X_0 - p_0 = \bigcup_{i=1}^{\infty} A_i, \tag{46}$$

we have $\dim X_0 \leqq n$ by the sum theorem (see [6], [10], [15]). On the other hand, f_0 is an essential mapping of X_0 into S_0^n and hence $\dim X_0 \geqq n$. Thus we have $\dim X_0 = n$.

Now we prove

$$\dim (X_0 \times Y) \geqq n+1 \tag{47}$$

for the following cases:

IV. X is compact and Y is a fully normal space of dimension $\geqq 1$.

V. X is fully normal and Y is the closed line interval $I = [0, 1]$.

Case IV. In this case X_0 is compact. Suppose, contrary to (47), that $\dim (X_0 \times Y) \leqq n$. Since $\dim Y \geqq 1$, there exist disjoint closed sets P, Q such that $\bar{V} - V \neq 0$ for any open set V satisfying $P \subset V \subset Y - Q$. Let us define a continuous mapping Φ_0 from $X_0 \times (P \cup Q)$ into S_0^n by

$$\Phi_0(x, y) = f_0(x) \quad \text{for} \quad x \in X_0, y \in P,$$

$$\Phi_0(x, y) = q_0 \quad \text{for} \quad x \in X_0, y \in Q.$$

Since $\dim (X_0 \times Y) \leqq n$, Φ_0 can be extended to a continuous mapping Φ from $X_0 \times Y$ into S_0^n by a well-known theorem ([3], [6], [10]). Let us put

$$\Phi_y(x) = \Phi(x, y)$$

and denote by V the set of points y of Y such that $\Phi_y : X_0 \rightarrow S_0^n$ is an essential mapping. Then we have obviously $P \subset V \subset Y - Q$. For each point y of Y there is an open neighbourhood $U(y)$ such that Φ_y and $\Phi_{y'}$ are homotopic for any point y' of $U(y)$, since X_0 is compact. Therefore V and $Y - V$ are open sets, that is, $\bar{V} - V = 0$. Thus we have arrived at a contradiction and (47) is established.

Case V. X_0 is fully normal and Y is the closed line interval I. Suppose that $\dim (X_0 \times I) \leqq n$. Then a continuous mapping Φ_0 of $(X_0 \times 0) \cup (X_0 \times 1)$ into S_0^n defined by

$$\Phi_0(x, 0) = f_0(x) \quad \text{for} \quad x \in X_0$$

$$\Phi_0(x, 1) = q_0 \quad \text{for} \quad x \in X_0$$

can be extended to a continuous mapping Φ of $X_0 \times Y$ into S_0^n. But the existence of such a mapping Φ means that the mapping f_0 is homotopic (in the ordinary sense; see [3, p. 204]) to a constant mapping. This contradicts the fact that f_0 is essential. This proves (47).

The relation (47) is thus established for both cases. The product space $X \times Y$ is normal and in view of (46)

$$X_0 \times Y = (p_0 \times Y) \cup (\bigcup_{i=1}^{\infty} (A_i \times Y)).$$

Since $p_0 \times Y$ is homeomorphic to a closed subset of $A_i \times Y$ for any non-empty A_i, by virtue of the sum theorem there exists some A_j $(j \geqq 1)$ such that $\dim (X_0 \times Y) = \dim (A_j \times Y)$. Hence we have, for Cases IV and V, $\dim (X \times Y) \geqq \dim (A_j \times Y) = \dim (X_0 \times Y) \geqq n+1$.

In case X is a locally compact fully normal space, there is a locally finite closed covering $\{F_\alpha\}$ of X such that F_α is compact, and hence by the generalized sum theorem [11, Theorem 3.2] we have $\dim X = \dim F_\alpha = n$ for some α. Therefore

$$\dim (X \times Y) \geqq \dim (F_\alpha \times Y) \geqq n+1,$$

where Y is a fully normal space of dimension $\geqq 1$. Thus we obtain

THEOREM 6. *If X is a locally compact fully normal space of dimension n $(n \geqq 0)$ and Y a fully normal space of dimension $\geqq 1$, then* $\dim (X \times Y) \geqq n+1$.

Our Theorem 6 clearly establishes the relation (B) for Case IV (see Introduction) in view of Theorem 4.

The discussion for Case V, together with Theorem 4, leads to the following theorem.

THEOREM 7. *If X is a fully normal space of dimension $\geqq 0$ and Y is a locally finite polytope of dimension $\geqq 0$, then*

$$\dim (X \times Y) = \dim X + \dim Y.$$

Finally we obtain from Theorems 4, 7 and Lemma 1:

THEOREM 8. *If X is a locally compact fully normal space of dimension $\geqq 0$ and Y is an arbitrary (finite or infinite) polytope of dimension $\geqq 0$ (see* 4), *then we have* $\dim (X \times Y) = \dim X + \dim Y$.

TOKYO UNIVERSITY OF EDUCATION,
TOKYO, JAPAN.

REFERENCES.

[1] E. G. Begle, "A note on S-spaces," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 577-579.

[2] J. Dieudonné, "Une généralisation des espaces compacts," *Journal de Mathematiques pures et appliquées*, vol. 23 (1944), pp. 65-76.

[3] C. H. Dowker, "Mapping theorems for non-compact spaces," *American Journal of Mathematics*, vol. 69 (1947), pp. 200-242.

[4] ———, "An extension of Alexandroff's mapping theorem," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 386-391.

[5] ———, "On countably paracompact spaces," *Canadian Journal of Mathematics*, vol. 3 (1951), pp. 219-224.

[6] E. Hemmingsen, "Some theorems in dimension theory for normal Hausdorff spaces," *Duke Mathematical Journal*, vol. 13 (1946), pp. 495-504.

[7] W. Hurewicz, "Sur la dimension des produits Cartésiens," *Annals of Mathematics* (2), vol. 36 (1935), pp. 194-197.

[8] S. Lefschetz, *Topics in Topology*, Princeton, 1942.

[9] K. Morita, "Star-finite coverings and the star-finite property," *Mathematica Japonicae*, vol. 1 (1948), pp. 60-68.

[10] ———, "On the dimension of normal spaces I," *Japanese Journal of Mathematics*, vol. 20 (1950), pp. 5-36.

[11] ———, "On the dimension of normal spaces II," *Journal of Mathematical Society of Japan*, vol. 2 (1950), pp. 16-33.

[12] R. H. Sorgenfrey, "On the topological product of paracompact spaces," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 631-632.

[13] E. H. Spanier, "Borsuk's cohomotopy groups," *Annals of Mathematics* (2), vol. 50 (1949), pp. 203-245.

[14] A. H. Stone, "Paracompactness and product spaces," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 677-688.

[15] A. D. Wallace, "Dimensional types," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 679-681.

[16] J. H. C. Whitehead, "Simplicial spaces, nuclei and m-groups," *Proceedings of the London Mathematical Society*, vol. 45 (1939), pp. 243-327.

COMPACTNESS CONDITIONS AND UNIFORM STRUCTURES.*

By Alice Dickinson.

In a completely regular topological space there exists at least one uniform structure compatible with the topology of the space. In this paper, relations between certain compactness conditions on a space and the set of compatible uniform structures on a space are considered. The terminology follows closely that of Bourbaki [2].

A uniform structure with filter $\mathfrak{F}_1$ is finer than a uniform structure with filter $\mathfrak{F}_2$ if $\mathfrak{F}_1$ is finer than $\mathfrak{F}_2$. With respect to this relation the uniform structures on a space form a partially ordered set. Weil [9], p. 16, showed that there is always one compatible uniform structure finer than all the others. This upper bound on the set of uniform structures is called the universal uniform structure of the space.

Consider the set of uniform structures compatible with the topology of a given space. Let a uniform structure in this set, less fine than all the other elements in the set, be called the *crude* uniform structure of the space.

THEOREM 1. *Let E be a locally compact space. Then the uniform structure induced by the uniquely defined compactification $E^* = E \cup \xi$ is the crude uniform structure of the space.*

Proof. Alexandroff and Urysohn [1], p. 68, proved that a locally compact space can be compactified by the addition of a single point in a unique manner. Let $\mathfrak{W}$ be a symmetric fundamental system of entourages of the uniform structure induced by this compact space, $E \cup \xi$. Let $\mathfrak{V}$ be a symmetric system of entourages of an arbitrary uniform structure compatible with the topology of E. Suppose the filter of $\mathfrak{V}$ is not finer than that of $\mathfrak{W}$. Then there exists an entourage W such that for every V_α in $\mathfrak{V}$, the set $V_\alpha - W$ is not empty. Let $F_\alpha = V_\alpha - W$. Then the sets $\{F_\alpha\}$ form the basis of a filter on $E \times E$. The filter basis $\{F_\alpha\}$ has no contiguous point in $E \times E$,

* Received August 20, 1952.

This paper is part of the author's doctoral dissertation submitted at the University of Michigan, 1952. Much of the dissertation was written while the author held a fellowship from the American Association of University Women.

since any such point would be contained both in $\bigcap_{\alpha} \bar{V}_\alpha = \Delta$ and in the complement of W; but these are disjoint sets. However, the filter basis $\{F_\alpha\}$ does have a contiguous point in the compact space $E^* \times E^*$. There are two cases to consider. First, suppose (ξ, ξ) is contiguous to the filter basis $\{F_\alpha\}$. The entourage W is the trace of an entourage W^* in $E^* \times E^*$. The entourage W^* contains an open set containing (ξ, ξ), and hence intersects every F_α. It follows that W intersects every F_α, which is contrary to the definition of the F_α. On the other hand, suppose the point (ξ, x_0) is contiguous to $\{F_\alpha\}$, and thus also (x_0, ξ) since the fundamental systems are symmetric. In E^* there exist disjoint neighborhoods $N_1(x_0)$ and $N_2(\xi)$. Since the uniform structure defined by $\mathfrak{V}$ is compatible with the topology of E, there is an entourage V_γ such that $V_\gamma(x_0) \subset N_1(x_0)$. And by the uniform structure axioms there is an entourage V_β such that $V_\beta \circ V_\beta \subset V_\gamma$. Let (r, s) be a point in $V_\beta \cap [V_\beta(x_0) \times N_2(\xi)]$ which is not empty since (x_0, ξ) is contiguous to $\{F_\alpha\}$. It follows that (x_0, r) is in V_β since r is in $V_\beta(x_0)$. Thus $(x_0, s) \in V_\beta \circ V_\beta \subset V_\gamma$ which implies that $s \in V_\gamma(x_0)$. Hence $V_\gamma(x_0) \cap N_2(\xi)$ is not empty. This is a contradiction.

Conversely, if the space E is not locally compact, then no uniform structure induced by a compactification is a crude uniform structure. In a space which is not locally compact, any compactification requires the addition of an infinite number of points. The identification of any two of these added points gives a compactification which induces a uniform structure which is less fine than that induced by the original compactification. This statement will be demonstrated by the construction in the proof of Theorem 2.

THEOREM 2. *If a space E has a unique uniform structure, then it has a unique compactification. This compactification consists of the addition of a single point. The unique uniform structure is, of course, induced by the compactification.*

Proof. Let $E^\dagger$ be a compactification obtained by the addition of points including $y_1 \neq y_2$. Let $\mathfrak{V}$ be a symmetric system of entourages defining the compatible uniform structure on $E^\dagger$. Let $\mathfrak{V}' = \{V'_\alpha\}$ be a new fundamental system of entourages, where

$$V'_\alpha = V_\alpha \cup [V_\alpha(y_1) \times V_\alpha(y_2)] \cup [V_\alpha(y_2) \times V_\alpha(y_1)]$$

for all $V_\alpha \in \mathfrak{V}$. Each V'_α is symmetric and contains $\Delta^\dagger$. Thus the satisfaction of the uniform structure axioms may be established by showing that

$V_\beta \circ V_\beta \subset V_\alpha$ implies $V'_\beta \circ V'_\beta \subset V'_\alpha$. If the points (a, b) and (b, d) are in V'_β, then one of the following combinations of elements, where $i \neq j$ take on the values 1 and 2, is in V_β: (1) $(a, b), (b, d)$; (2) $(y_i, a), (y_j, b), (y_i, d)$; (3) $(y_i, a), (y_j, b), (b, d)$; (4) $(a, b), (y_i, b), (y_j, d)$; (5) $(y_i, a), (y_j, d)$. If $V_\beta \circ V_\beta \subset V_\alpha$, (1) and (2) imply that (a, d) is in V_α, while (3), (4), and (5) imply that (y_i, a) and (y_j, d) are in V_α. In either case (a, d) is in V'_α. Hence $V'_\beta \circ V'_\beta \subset V'_\alpha$ and the $\{V'_\alpha\}$ define a uniform structure. Although this new uniform structure is not compatible with the topology of $E^\dagger$, its trace on E is useful. For any point x in E, there is a $V_\gamma \varepsilon \mathfrak{V}$ such that $V_\gamma(x)$ does not contain y_1 or y_2. Then $V'_\gamma(x) \subset V_\gamma(x)$. Hence the uniform structures on E defined by the traces of $\mathfrak{V}$ and $\mathfrak{V}'$ on $E \times E$ are both compatible with the topology of E. Since $\mathfrak{V}$ defines the filter of all neighborhoods of $\Delta^\dagger$, there is an entourage V_μ and a neighborhood $W(y_1, y_2)$ in $E^\dagger \times E^\dagger$ which are disjoint. The trace of V_μ on $E \times E$ does not contain the trace of any V'_α. Thus the two uniform structures induced on E are distinct.

THEOREM 3. *If a space has a unique uniform structure it is countably compact.*

Proof. If the space is not countably compact there exists an ordered set $\{x_n\}$ of distinct points without a limit point. Since the space is completely regular there exists an open set $W(x_1)$ such that $\overline{W(x_1)} \cap \bigcup_{j=2}^{\infty} x_j = 0$. For each $n \geqq 2$, there exists an open set $W'(x_n)$ such that $\overline{W'(x_n)} \cap \bigcup_{j=n+1}^{\infty} x_j = 0$. Let $W(x_n)$ be an open set in the intersection of $W'(x_n)$ and the complement of $\bigcup_{j=1}^{n-1} \overline{W(x_j)}$. Then the $\bigcup_{n=1}^{\infty} W(x_{2n})$ and $\bigcup_{n=1}^{\infty} W(x_{2n+1})$ form disjoint open sets which separate two closed sets, $\{x_{2n}\}$ and $\{x_{2n+1}\}$, neither of which is compact. However, according to the condition given by Doss [5], this contradicts the hypothesis.

The converse is not true. Consider a space consisting of two distinct sets of the ordinal numbers of the first and second classes with the usual topology on each set. This space is countably compact but does not have a unique uniform structure.

A space is compact if and only if it has a unique uniform structure relative to which it is complete. The question of determining which topological spaces (weaker than compact spaces) always admit a uniform structure

relative to which they are complete, was raised by André Weil [9], p. 38. Dieudonné [4] showed that normality was not sufficient. In his paper on paracompact spaces [3] he posed the problem of whether or not the universal uniform structure of a paracompact space is the uniform structure of a complete space. The affirmative answer is given here.

THEOREM 4. *A paracompact space E is complete with respect to its universal uniform structure.*

Proof. Suppose the universal uniform structure of E is not complete. Then there exists a Cauchy filter $\mathfrak{F}$ which does not converge; that is, for every x in E, there exists an open set $W(x)$ which does not contain any F_α in $\mathfrak{F}$. Consider the covering $\mathfrak{W}$ consisting of all the open sets $W(x)$ for every x in E. Since E is paracompact, there exists a covering $\mathfrak{U} = \{U_\alpha\}$ such that, for every x in E, the star $(x, \mathfrak{U}) = \bigcup_{x \in U_\alpha} U_\alpha$ is contained in some set of $\mathfrak{W}$ [6]. Then the open set $U(\Delta) = \bigcup_\alpha (U_\alpha \times U_\alpha)$ is an entourage of the universal uniform structure of E. Since $\mathfrak{F}$ is a Cauchy filter, there is a set F_0 in $\mathfrak{F}$ such that $F_0 \times F_0 \subset U(\Delta)$. Let p be any fixed point in F_0. For every point y in F_0 there is a covering set U_β containing p and y. Since $(p, y) \in (F_0 \times F_0) \subset U(\Delta)$, the point (p, y) must be an element of some set $U_\beta \times U_\beta$. Thus $F_0 \subset$ star $(p, \mathfrak{U}) \subset W(x)$ for some x, which is absurd.

This gives as immediate corollaries some known [8] relations.

COROLLARY 1. *If a paracompact space is precompact in all its uniform structures it is compact.*

COROLLARY 2. *A paracompact space with a unique uniform structure is compact.*

A converse statement for Theorem 4 poses an interesting problem.

UNIVERSITY OF MICHIGAN,
PENNSYLVANIA STATE COLLEGE.

BIBLIOGRAPHY.

[1] Alexandroff, P. and Urysohn, P., "Mémoire sur les espaces topologiques compacts," *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam*, vol. 14 (1929), No. 1.

[2] Bourbaki, N., *Éléments de Mathématique*, Livre III. *Topologie Générale*, Actualités Scientifiques et Industrielles, no. 858, Paris: Hermann, 1940.

[3] Dieudonné, J., "Une généralisation des espaces compacts," *Journal de Mathématiques Pures et Appliquées*, vol. 23 (1944), pp. 65-76.

[4] ———, "Exemple d'espace normal non susceptible d'une structure uniform d'espace complet," *Comptes Rendus, Hebdomadaires des Seances de l'Academie des Sciences*, vol. 209 (1939), pp. 145-147.

[5] Doss, R., "On uniform spaces with a unique structure," *American Journal of Mathematics*, vol. 71 (1949), pp. 19-23.

[6] Stone, A. H., "Paracompactness and product spaces," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 977-982.

[7] Tukey, J. W., *Convergence and Uniformity in General Topology*, Annals of Mathematics Studies, no. 2, Princeton, 1940.

[8] Umegaki, H., "On the uniform space," *Tohoku Mathematical Journal*, Second Series, vol. 2, No. 1 (1950), pp. 57-63.

[9] Weil, A., *Sur les espaces a structure uniforme et sur la topologie générale*, Actualités Scientifiques et Industrielles, no. 551, Paris: Hermann, 1937.

ON THE ESSENTIAL SPECTRA OF SYMMETRIC OPERATORS IN HILBERT SPACE.*

By PHILIP HARTMAN.

The object of this paper is to obtain some results concerning the relationships between the spectra of the self-adjoint extensions of a closed, symmetric operator T on Hilbert space (having equal deficiency indices). These results are known in the case of certain second order, ordinary differential operators T. The proofs in these cases, however, usually depend on functions and operations having no analogues in Hilbert space and, consequently, are not valid in the general case.

1. Point spectra. Let $\mathfrak{H}$ denote a Hilbert space; T a closed, symmetric operator in $\mathfrak{H}$ with domain $\mathfrak{D}(T)$, range $\mathfrak{R}(T)$, adjoint T^* and the deficiency indices (m, m). If l is a complex number, $\mathfrak{M}(l)$ will denote the manifold of elements x of $\mathfrak{H}$ satisfying $(T^* - lI)x = 0$. If $\mathfrak{I}(l) \neq 0$, then the manifolds $\mathfrak{R}(T - lI)$ and $\mathfrak{M}(\bar{l})$ are closed, orthogonal and span $\mathfrak{H}$. In addition, $\mathfrak{D}(T)$, $\mathfrak{M}(l)$, $\mathfrak{M}(\bar{l})$ are in $\mathfrak{D}(T^*)$ and every element x in $\mathfrak{D}(T^*)$ has a unique representation of the form

$$x = x_0 + x(l) + x(\bar{l}), \qquad \mathfrak{I}(l) \neq 0, \tag{1}$$

where x_0, $x(l)$, $x(\bar{l})$ are in $\mathfrak{D}(T)$, $\mathfrak{M}(l)$, $\mathfrak{M}(\bar{l})$, respectively.

For a fixed non-real l, there is a one-to-one correspondence between the self-adjoint extensions A of T and the norm-preserving (linear, continuous) transformations $V = V(l)$ of $\mathfrak{M}(l)$ onto $\mathfrak{M}(\bar{l})$ such that if $A \rightarrow V$, then $\mathfrak{D}(A)$ consists of those elements of the form

$$x = x_0 + x(l) + Vx(l), \qquad \mathfrak{I}(l) \neq 0, \tag{2}$$

where x_0, $x(l)$ are arbitrary elements of $\mathfrak{D}(T)$, $\mathfrak{M}(l)$, respectively (and $Vx(l)$ is in $\mathfrak{M}(\bar{l})$). Cf., e. g., [12], Chap. IX or [9], Chap. VI.

(α) *Let x be an element of $\mathfrak{D}(T^*)$ with the representation* (1) *and satisfying, for some real number λ,*

$$(T^* - \lambda I)x = 0. \tag{3}$$

* Received September 3, 1952.

Then

$$\| x(l) \| = \| x(\bar{l}) \|. \tag{4}$$

Proof. Without loss of generality it can be supposed that $\lambda = 0$. According to (1) and (3), $-lx = (T^* - lI)x = (T - lI)x_0 - 2i\mathfrak{J}(l)x(\bar{l})$. Hence $\| lx \|^2 = \| (T - lI)x_0 \|^2 + (2\mathfrak{J}(l))^2 \| x(\bar{l}) \|^2$. Similarly,

$$\| lx \|^2 = \| (T - \bar{l}I)x_0 \|^2 + (2\mathfrak{J}(l))^2 \| x(l) \|^2.$$

Since the symmetry of T implies $\| (T - lI)x_0 \| = \| (T - \bar{l}I)x_0 \|$, the result (4) follows.

It is clear from the relationship of $\mathfrak{D}(A)$ to $\mathfrak{M}(l)$ that (α) implies the following assertion:

Corollary. *If x in $\mathfrak{D}(T^*)$ satisfies (3), then there exist self-adjoint extensions A of T for which x is in $\mathfrak{D}(A)$, (and, hence, satisfies $(A - \lambda I)x = 0$).*

The next lemma corresponds to a part of Weyl's theorem [14], p. 238, which states that, in his case of differential operators, the deficiency index m can be determined by the consideration of $T^* - \lambda I$ for real λ; cf. (γ) below.

(β) *If the dimension of the manifold of elements x in $\mathfrak{D}(T)$ satisfying $(T - \lambda I)x = 0$ is π, where $0 \leqq \pi \leqq \infty$, then there are at most $m + \pi$ linearly independent solutions of (3); so that, if λ is in the point spectrum of a self-adjoint extension A of T, its multiplicity is at least π and at most $m + \pi$.*

Proof. It can be supposed that π and m are finite, for otherwise (β) is trivial. Let $m < \infty$ and $\pi = 0$ (the proof for $\pi > 0$ is similar). Suppose that x satisfies (3) and has the representation (1) for some non-real l. Then $x(l) \neq 0$, for otherwise $x(\bar{l}) = 0$, by (α), and x is in $\mathfrak{D}(T)$, which contradicts the hypothesis, $\pi = 0$. Since there are only m linearly independent elements $x(l)$, it follows that if there were $m + 1$ linearly independent solutions of (3), then there would exist a linear combination of them, say x, for which the corresponding $x(l)$ is 0. This proves (β).

The following assertion is an analogue of results on differential operators; e. g., Weyl [14], pp. 221-227 and Hartman and Wintner, [1]. The proof is adapted from the proof in Stone [12], pp. 489-491, of Weyl's theorem in [14], p. 238.

(γ) *Let A be a self-adjoint extension of T and λ a complex, possibly real, number in the resolvent set of A, then there exist exactly m linearly independent elements x in $\mathfrak{D}(T^*)$ satisfying (3) (and no such element is in $\mathfrak{D}(T)$).*

Proof. Only the case of a real λ has to be considered. Let g be an element of the manifold $\mathfrak{M}(l_0)$, where $\mathfrak{J}(l_0) \neq 0$. Let $w = (\lambda - l_0)^{-1}$. Then w is in the resolvent set of the bounded, normal operator $(A - l_0 I)^{-1}$ and, therefore, the equation

$$(5) \qquad \{(A - l_0 I)^{-1} - wI\}x = g$$

has a unique solution $x = x_g$. This equation can be written as $(A - l_0 I)^{-1}x = wx + g$ or $x = (A - l_0 I)(wx + g)$. Since $wx + g$ is in $\mathfrak{D}(A)$ (hence, in $\mathfrak{D}(T^*)$) and g is in $\mathfrak{D}(T^*)$, it follows that x is in $\mathfrak{D}(T^*)$. Thus the last equation gives $x = (T^* - l_0 I)wx$ or (3), since $w = (\lambda - l_0)^{-1} \neq 0$. Thus, to every g in $\mathfrak{M}(l_0)$, there corresponds an $x = x_g$ of $\mathfrak{D}(T^*)$ satisfying (3). It is clear from (5) that to linearly independent g there correspond linearly independent x_g. Hence (3) has at least m linearly independent solutions x_g.

No element $x \neq 0$ satisfying (3) is in $\mathfrak{D}(T)$, for otherwise $(A - \lambda I)x = (T - \lambda I)x = 0$ contradicts the fact that λ is in the resolvent set of A. Thus (γ) follows from (β).

Remark. The reality of λ was used in the proof above only to assure that $\lambda - l_0 \neq 0$. Thus λ, l_0 can be any pair of numbers in the resolvent set of A with the restriction $\mathfrak{J}(l_0) \neq 0$ and $\lambda \neq l_0$. Writing l in place of λ in w, it follows that $x = x_g(l)$, for a fixed g, depends analytically on l (in the sense that for a fixed element y of $\mathfrak{H}$, the scalar product $(x_g(l), y)$ is a regular analytic function of l). Clearly, there exists a set of m orthonormal elements $x^1(l), x^2(l), \cdots$ spanning $\mathfrak{M}(l)$ and depending continuously on l. Hence, if E_l is the projection on $\mathfrak{M}(l)$, then, for a fixed y, $E_l y$ depends continuously on l on the resolvent set of any self-adjoint extension A of T (in particular, on the half-planes $\mathfrak{J}(l) \neq 0$). Cf. [3], [6].

(δ) *Let A_1, A_2 be self-adjoint extensions of T and let l belong to the resolvent set of both A_1, A_2 (for example, let $\mathfrak{J}(l) \neq 0$) and let $D = D(l) = (A_2 - lI)^{-1} - (A_1 - lI)^{-1}$. Then $D = E_l D = D E_{\bar{l}}$.*

In other words, to a complete orthonormal set $x^1, x^2, \cdots$ in $\mathfrak{M}(l)$, there corresponds a set of elements $h_1, h_2, \cdots$ in $\mathfrak{M}(\bar{l})$ such that

$$(6) \qquad Dy = (y, h_1)x^1 + (y, h_2)x^2 + \cdots.$$

In particular, if $m < \infty$, then D is completely continuous.

(δ) and the last remark are generalizations of the result of Weyl [14], p. 251, deduced from the Green kernel representations of $(A_1 - lI)^{-1}$,

$(A_2 - lI)^{-1}$, when $\mathfrak{I}(l) \neq 0$; cf. [2], pp. 314-315 or [6], p. 782, for the corresponding case of a real l.

A modification of the proof of (δ) shows that if $\mathfrak{I}(l) \neq 0$ and V_1, V_2 are the isometric mappings of $\mathfrak{M}(l)$ onto $\mathfrak{M}(\bar{l})$ associated with A_1, A_2, respectively, then $-2\mathfrak{I}(l)Dy = (V_2^{-1} - V_1^{-1})E_l y$.

(ϵ) *If l is real in (δ), then D is bounded, self-adjoint and commutes with E_l. Conversely, if l is real and in the resolvent set of A_1 and D is a bounded, self-adjoint operator which commutes with E_l, then $(A_1 - lI)^{-1} + D$ is the inverse of a self-adjoint operator, say $A_2 - lI$, where A_2 is an extension of T.*

If l is real and $m = 1$, then $D = cE_l$, where c is an arbitrary real constant; for the ordinary differential operator case, see [6], p. 782.

Proof of (δ). Let y be any element of $\mathfrak{H}$ and put $(A_j - lI)^{-1}y = x_j$. Then x_j is in $\mathfrak{D}(A_j - lI)$, hence in $\mathfrak{D}(T^*)$; so that $y = (T^* - lI)x_j$. On letting $j = 1, 2$ and subtracting these relations, it is seen that

$$0 = (T^* - lI)(x_2 - x_1).$$

In other words, $Dy = x_2 - x_1$ is in $\mathfrak{M}(l)$. Hence $D = E_l D$.

In order to prove $D = DE_{\bar{l}}$, it is sufficient to verify that $Dy = 0$ whenever y is orthogonal to $\mathfrak{M}(\bar{l})$. If y is such an element and $\mathfrak{I}(l) \neq 0$, then y is in the range, $\mathfrak{R}(T - lI)$, of $T - lI$, say $y = (T - lI)x$. It follows that $(A_j - lI)^{-1}y = x$ for $j = 1, 2$, and, consequently, $Dy = 0$. If l is real, then y is in the closure of the linear manifold $\mathfrak{R}(T - lI)$, and $Dy = 0$ follows from continuity considerations. This proves (δ).

Proof of (ϵ). If l is real, D is the difference of two bounded, self-adjoint operators; so that the first assertion of (ϵ) is trivial, since $l = \bar{l}$.

Let l and D satisfy the assumptions of the last part of (ϵ). It can be supposed that $l = 0$. Define an operator A_2 by $A_2 = T^*$, where $\mathfrak{D}(A_2)$ consists of those elements x representable in the form $x = (A_1^{-1} + D)y$ for some y in $\mathfrak{H}$. An element x has at most one such representation. In order to see this, note that $x = (A_1^{-1} + D)y$ and $x = (A_1^{-1} + D)z$ imply that $(x - Dy) - (x - Dz) = A_1^{-1}(z - y)$ is in $\mathfrak{D}(A_1)$. Since $D = E_0 D$, $A_1 D(z - y) = T^* D(z - y) = 0$; so that $D(z - y) = 0$ since A_1 is non-singular. Thus $0 = x - x = A_1^{-1}(z - y)$ implies that $z - y = 0$. It is clear that $\mathfrak{D}(A_2)$ contains $\mathfrak{D}(T)$ and is contained in $\mathfrak{D}(T^*)$; so that the definition $A_2 = T^*$ in $\mathfrak{D}(A_2)$ is meaningful. The range of A_2 is $\mathfrak{H}$; thus, if it is

shown that A_2 is symmetric, then (ϵ) will be proved. To this end, let $u=(A_1^{-1}+D)v$, $x=(A_1^{-1}+D)y$ be two elements of $\mathfrak{D}(A_2)$. Then $(A_2u, x)=(v, (A_1^{-1}+D)y)=(u, y)=(u, A_2x)$. This proves ($\epsilon$).

A slight generalization of (ϵ) is given by

(ϵ'). *Let B be a self-adjoint operator such that $Bx=0$ implies $x=0$. Let $\mathfrak{M}$ be a (closed) m-dimensional manifold, $0<m\leqq\infty$, with the property that the equation $Bx=y$ has no solution whenever $y\neq 0$ is in $\mathfrak{M}$. Let $\mathfrak{D}$ be the set of elements x representable in the form $x=Bz$, where z is orthogonal to $\mathfrak{M}$, and let T be the operator, with domain $\mathfrak{D}(T)=\mathfrak{D}$, defined by $Tx=z$ when $x=Bz$. Then T is a closed, symmetric operator with deficiency indices (m, m) and B^{-1} is a self-adjoint extension of T. $\mathfrak{D}(T^*)$ is the set of elements of the form $x+y$, where x is in $\mathfrak{R}(B)$, say $x=Bz$, and y is in $\mathfrak{M}$; finally, $T^*(x+y)=z$.*

The proof is straightforward and will be omitted.

2. A lemma. If A is a self-adjoint operator, let $\Sigma(A)$ denote its spectrum, $\Pi(A)$ its point spectrum, $\Sigma'(A)$ its essential spectrum and $\Gamma(A)$ its continuous spectrum. By the essential spectrum of A is meant the set of finite cluster points of $\Sigma(A)$, including the points in the point spectrum of infinite multiplicity. Thus $\Sigma(A)=\Pi(A)+\Sigma'(A)$, although $\Pi(A)$ and $\Sigma'(A)$ need not be disjoint. By the continuous spectrum of A is meant the set of λ for which the multipilicty of λ in the continuous spectrum is at least 1; cf. [12], p. 267. Thus $\Sigma'(A)$ contains $\Gamma(A)$.

Let X be a closed operator in Hilbert space $\mathfrak{H}$ with a domain $\mathfrak{D}(X)$, dense in $\mathfrak{H}$. It is not supposed that X is symmetric. The object of this section is to point out the relationships between the spectra of the (non-negative) self-adjoint operators X^*X and XX^*. Such relationships are of interest in view of Toeplitz's remarks on the existence of "left" and "right" inverses of X; cf. [16], pp. 138-139, and will have applications below. The results will be based on a theorem of Neumann [10], p. 307, which, in turn, is a generalization of the result of Wintner [17], p. 145 (cf. [15], p. 282), that if X is non-singular (and bounded), then X can be represented as a product of a positive-definite, self-adjoint operator and a unitary operator.

LEMMA. *Let X be a closed operator with a domain $\mathfrak{D}(X)$ which is dense in $\mathfrak{H}$. Let $E(\lambda)$, $F(\lambda)$ be the resolutions of the identity belonging to X^*X, XX^*, respectively (where $E(\lambda)$, $F(\lambda)$ are continuous from the right). Then there exists a unique bounded operator W with the properties that*

$$W^*W=I-E(0) \text{ and } WW^*=I-F(0) \tag{7}$$

and that if $0 \leqq \lambda \leqq \mu$, *then*

$$(8)\quad F(\mu) - F(\lambda) = W\{E(\mu) - E(\lambda)\}W^* \text{ and } E(\mu) - E(\lambda) = W^*\{F(\mu) - F(\lambda)\}W.$$

These relations imply that if $y = Wx$ (or if $x = W^*y$) and $0 \leqq \lambda < \mu$, then $\|\{E(\mu) - E(\lambda)\}x\| = \|\{F(\mu) - F(\lambda)\}y\|$. Hence, every $\lambda \neq 0$ has the same multiplicity in the point spectra $\Pi(X^*X)$ and $\Pi(XX^*)$ and every λ has the same multiplicity in the continuous spectra $\Gamma(X^*X)$ and $\Gamma(XX^*)$. The exceptional standing of $\lambda = 0$ arises, of course, from the fact that although $E(-0) = F(-0) = 0$, the projections $E(0)$, $F(0)$ need not be 0 and, in fact, their ranges can be manifolds of different dimensionality. Thus $\lambda = 0$ can be in one of the sets $\Pi(X^*X)$, $\Pi(XX^*)$ without being in the other or, more generally, the multiplicities of $\lambda = 0$ in the sets $\Pi(X^*X)$, $\Pi(XX^*)$ can be different. Except for this possibility, the spectra (counting multiplicities) of X^*X and XX^* are identical.

It is clear from the Lemma (without any appeal to the Hellinger theory [8] or its extensions) that X^*X and XX^* are unitarily equivalent if and only if the multiplicities of $\lambda = 0$ in $\Pi(X^*X)$ and $\Pi(XX^*)$ are equal. For if these multiplicities are equal, the definition of W (which is 0) on the range of $E(0)$ can be altered so that W gives an isometric mapping of the range of $E(0)$ onto the range of $F(0)$. The altered W, say U, will be unitary and will satisfy $XX^* = U^*X^*XU$.

In this case (when the multiplicities of $\lambda = 0$ in $\Pi(X^*X)$ and $\Pi(XX^*)$ are equal), the theorem of Wintner mentioned above has the extension that X is the product of a non-negative self-adjoint operator and of a unitary operator, in fact, $X = (XX^*)^{\frac{1}{2}}U = U^*(X^*X)^{\frac{1}{2}}$.

The Lemma and its consequences were suggested to me by a discussion with Professor Wintner of the Appendix of [11], pp. 75-78.

Proof of the Lemma. According to [10], p. 307, there exists a (unique) bounded operator W with the properties that it maps $\mathfrak{R}((X^*X)^{\frac{1}{2}})$ isometrically onto $\mathfrak{R}((XX^*)^{\frac{1}{2}})$ and satisfies (7) and $X = W(X^*X)^{\frac{1}{2}} = (XX^*)^{\frac{1}{2}}W^*$. Furthermore, for every x in $\mathfrak{H}$, the element Wx is in the closed linear manifold spanned by $\mathfrak{R}((XX^*)^{\frac{1}{2}})$ and, hence, is orthogonal to the range of $F(0)$ (that is, to the elements satisfying $(XX^*)^{\frac{1}{2}}y = 0$ or equivalently $X^*y = 0$). Thus $F(0)W = 0$. From (7), it is clear that $W^*F(0) = 0$. These two relations, their analogues $E(0)W^* = WE(0) = 0$ and the relations (7) make it clear that if $G(\lambda)$ is defined to be 0 or $F(0) + WE(\lambda)W^*$ according as $\lambda < 0$ or $\lambda \geqq 0$, then $G(\lambda)$ is a resolution of the identity. It follows from

$XX^* = WX^*XW^*$ that $G(\lambda) = F(\lambda)$, that is, that $F(\lambda)$ is 0 or $F(0) + WE(\lambda)W^*$ according as $\lambda < 0$ or $\lambda \geqq 0$. This implies (8) and completes the proof of the Lemma.

Remark. The theorem of Neumann, referred to above, has also the following consequences: For a given y, the existence of a solution x or z for one of the equations

$$X^*x = y \quad \text{and} \quad (X^*X)^{\frac{1}{2}}z = y$$

implies the existence of a solution z or x for the other equation. This follows from $X^* = (X^*X)^{\frac{1}{2}}W^*$. For if x is a solution for the first equation then $z = W^*x$ is a solution for the second. If z is the solution of least norm of the second equation, so that $(I - E(0))z = z$, then $x = Wz$ satisfies $W^*x = W^*Wz = z$, by (7), and is a solution of the first equation. In other words, $\mathfrak{R}((X^*X)^{\frac{1}{2}}) = \mathfrak{R}(X^*)$ (this contrasts with the relation $\mathfrak{D}((X^*X)^{\frac{1}{2}}) = \mathfrak{D}(X)$; [10], p. 304). Thus $X^*x = y$ fails to have a solution for some y if and only if $\lambda = 0$ is in the spectrum $\Sigma((X^*X)^{\frac{1}{2}})$, that is, in $\Sigma(X^*X)$.

3. Essential spectra. As in Section **1**, T will denote a closed, symmetric operator with deficiency indices (m, m). Let the set of (real) λ for which there is an $x \neq 0$ in $\mathfrak{H}$ satisfying

$$(9) \qquad (T - \lambda I)x = 0,$$

be called $\Pi(T)$. Correspondingly, let the set of *real* λ for which there is an $x \neq 0$ satisfying (3) be called $\Pi_0(T^*)$. Let $\Sigma_0(T^*)$ denote the set of *real* λ to which there corresponds at least one element y in H with the property that

$$(10) \qquad (T^* - \lambda I)x = y$$

has no solution x.

It can be remarked that $\lambda = \mu$ is in $\Pi(T)$ if and only if $\lambda = 0$ is in $\Pi((T^* - \mu I)(T - \mu I))$, the point spectrum of the self-adjoint operator $(T^* - \mu I)(T - \mu I)$. Similarly, $\lambda = \mu$ is in $\Pi_0(T^*)$ if and only if $\lambda = 0$ is in $\Pi((T - \mu I)(T^* - \mu I))$. In view of the Remark at the end of the last section, $\lambda = \mu$ is in $\Sigma_0(T^*)$ if and only if $\lambda = 0$ is in $\Sigma((T^* - \mu I)(T - \mu I))$.

Let Σ_1' and Σ_2' denote the sets of μ-values for which $\lambda = 0$ is in $\Sigma'((T^* - \mu I)(T - \mu I))$ and $\Sigma'((T - \mu I)(T^* - \mu I))$, respectively. Since $\Pi(T)$ is contained in $\Pi_0(T^*)$, it follows from the Lemma of the last section that Σ_1' is contained in Σ_2'. Furthermore, if $m < \infty$, then $\Sigma_1' = \Sigma_2'$ by virtue of (β), since a necessary condition for a $\lambda = \mu$ in Σ_2' to fail to belong to Σ_1'

is that $\lambda = \mu$ be in $\Pi(T)$ with a finite, and in $\Pi_0(T^*)$ with an infinite multiplicity.

In what follows, A_1, A_2 and A represent self-adjoint extensions of A.

(i) *If $0 < m < \infty$ and if the interval $\mu \leqq \lambda \leqq \nu$ contains $m+1$ points of the spectrum of A_1 (where points of the point spectrum are counted with their multiplicities), then the same interval contains at least one point of the spectrum of A_2. In particular, $\Sigma'(A_1) = \Sigma'(A_2)$.*

The first part of this assertion is clear from (δ). The last part is a consequence of the first part; in this assertion, "finite cluster points" in the definition of the essential spectrum $\Sigma'(A)$ can be replaced by "finite and infinite cluster points." (This last remark does not follow if one does not use the full force of (δ), but only its consequence that D is completely continuous.) Thus if A_1 is bounded (or unbounded) from above, so is A_2.

The last part of the italicized assertion above is a generalization of a result of Weyl [14], p. 251, for differential operators. The generalization (i) to arbitrary symmetric operators of the type here considered is due to E. Heinz [7]. Other separation theorems for the case $m = 1$ are given in [4], [6].

(ii) *A point $\lambda = \mu$ is in at least one $\Pi(A)$ if and only if it is in $\Pi_0(T^*)$, that is, if and only if $\lambda = 0$ is in $\Pi((T-\mu I)(T^*-\mu I))$. A point $\lambda = \mu$ is in every $\Pi(A)$ if and only if it is in $\Pi(T)$, that is, if and only if $\lambda = 0$ is in $\Pi((T^*-\mu I)(T-\mu I))$.*

The situation with respect to essential spectra is somewhat more complicated, when $m = \infty$:

(iii) *If a point $\lambda = \mu$ is in at least one $\Sigma'(A)$, then it is in Σ_2', that is, $\lambda = 0$ is in $\Sigma'((T-\mu I)(T^*-\mu I))$. If a point $\lambda = \mu$ is in Σ_1', that is, if $\lambda = 0$ is in $\Sigma'((T^*-\mu I)(T-\mu I))$, then it is in every $\Sigma'(A)$.*

When $0 < m < \infty$, then $\Sigma'(A)$ is independent of A and is identical with the set $\Sigma_1' = \Sigma_2'$.

(iv) *Every point $\lambda = \mu$ is in at least one $\Sigma(A)$ and, for every μ, the point $\lambda = 0$ is in $\Sigma((T-\mu I)(T^*-\mu I))$. If a point $\lambda = \mu$ is in $\Sigma_0(T^*)$, that is, if $\lambda = 0$ is in $\Sigma((T^*-\mu I)(T-\mu I))$, then $\lambda = \mu$ is in every $\Sigma(A)$; conversely, when $0 < m < \infty$, if a point $\lambda = \mu$ is in $\Sigma(A)$ for every A, then it is in $\Sigma_0(T^*)$.*

These assertions are known for certain ordinary differential operators, where $m = 1$; Hartman and Wintner [5].

Proof of (ii). Since T^* is an extension of A, it is clear that $\Pi_0(T^*)$ contains $\Pi(A)$. Conversely, if $\lambda = \mu$ is in $\Pi_0(T^*)$, then it is in some $\Pi(A)$ by the Corollary of (α). This proves the first assertion of (ii).

Clearly, if $\lambda = \mu$ is in $\Pi(T)$, then it is in every $\Pi(A)$. In order to see that if $\lambda = \mu$ is not in $\Pi(T)$, then there is an A for which $\lambda = \mu$ is not in $\Pi(A)$, let $x^1, x^2, \cdots$ be a complete set of linearly independent elements in $\mathfrak{M}(\mu)$. For some non-real l, let $x^j = x_0{}^j + x^j(l) + x^j(\bar{l})$ be the decomposition (1) of x^j. Suppose $x^1, x^2, \cdots$ have been selected so that $x^1(l)$, $x^2(l), \cdots$ form an orthonormal sequence. Then, by (α), $x^1(\bar{l}), x^2(\bar{l}), \cdots$ is also an orthonormal sequence. Consider an isometric mapping V_1 of $\mathfrak{M}(l)$ onto $\mathfrak{M}(\bar{l})$, for which $V_1 x^j(l) = x^j(\bar{l})$. Then any isometric mapping V of $\mathfrak{M}(l)$ onto $\mathfrak{M}(\bar{l})$ can be represented as $V = UV_1$, where U is a unitary mapping of $\mathfrak{M}(\bar{l})$ onto $\mathfrak{M}(\bar{l})$. Let U be chosen so that $\lambda = 1$ is not in its point spectrum. Then it readily follows that no element $x \neq 0$ in $\mathfrak{M}(\mu)$ is in $\mathfrak{D}(A)$, where A is the self-adjoint extension of T corresponding to V. Thus $\lambda = \mu$ is not in $\Pi(A)$.

Proof of (iii). The proof will depend on the following criterion of Weyl, [13]: For any self-adjoint operator B, $\lambda = 0$ is in $\Sigma'(B)$ if and only if there exists a sequence $x_1, x_2, \cdots$ in $\mathfrak{D}(B)$ such that $\| x_n \| = 1$, $x_n \to 0$ (weakly) and $Bx_n \to 0$ (strongly), as $n \to \infty$.

It can be supposed that $\mu = 0$. The second assertion of (iii) will be proved first. Suppose that $\lambda = 0$ is in Σ_1', that is, in $\Sigma'(T^*T)$. Then there is a sequence $x_1, x_2, \cdots$ with the properties described above, where $B = T^*T$. Since $\| x_n \| = 1$ and $T^*Tx_n \to 0$ (strongly), as $n \to \infty$, it follows that $(T^*Tx_n, x_n) \to 0$ as $n \to \infty$; that is, $Tx_n \to 0$ (strongly) as $n \to \infty$. Since every A is an extension of T, it follows that $Ax_n \to 0$ (strongly), as $n \to \infty$. Thus $\lambda = 0$ is in $\Sigma'(A)$ for every A.

The proof of the first assertion of (iii) is proved similarly. Let $\mu = 0$ and let $x_1, x_2, \cdots$ have the properties described in Weyl's criterion, where $B = A$ for some A. Then x_n, which is in $\mathfrak{D}(A)$, is in $\mathfrak{D}(T^*)$ and satisfies $T^*x_n = Ax_n$. Hence $T^*x_n \to 0$ (strongly), which implies that $(TT^*)^{\frac{1}{2}}x_n \to 0$ (strongly); cf. [10], p. 304. Thus $\lambda = 0$ is in $\Sigma'((TT^*)^{\frac{1}{2}})$ and, hence, in $\Sigma'(TT^*)$. Consequently, $\lambda = 0$ is in Σ_2', as was to be proved.

Proof of (iv). If $\lambda = \mu$ is not in $\Sigma(A)$, for some A, then $\lambda = \mu$ is in $\Pi_0(T^*)$, by (γ). Hence $\lambda = \mu$ is in $\Pi(A_1)$ for some A_1, by (α). This proves the first part of the first assertion in (iv).

In order to prove the second part, note that if $\lambda = 0$ is not in $\Pi((T-\mu I)(T^*-\mu I))$, then $\lambda = \mu$ is in $\Sigma(A)$ for every A, by (γ), and hence, is in $\Sigma'(A)$, by (ii). Consequently, (iii) implies that $\lambda = 0$ is in $\Sigma'((T-\mu I)(T^*-\mu I))$; in others words, for every μ, $\lambda = 0$ is in

$$\Pi((T-\mu I)(T^*-\mu I)) + \Sigma'((T-\mu I)(T^*-\mu I)) = \Sigma((T-\mu I)(T^*-\mu I)).$$

There remains to prove the second assertion of (iv), that concerning $\Sigma_0(T^*)$. The last parts of (ii) and (iii) show that if $\lambda = 0$ is in $\Sigma((T^*-\mu I)(T-\mu I))$, then $\lambda = \mu$ is in $\Sigma(A)$ for every A. When $m < \infty$, the converse is true, since $m < \infty$ implies $\Sigma_1' = \Sigma_2'$. Hence (iv) follows from the Remark at the end of the last section (if X is identified with $T - \mu I$).

4. Continuous spectra. For a self-adjoint operator A and a real number λ, let $p(\lambda) = p(\lambda, A)$ and $c(\lambda) = c(\lambda, A)$ denote the multiplicities of λ in the point and continuous spectra of A, respectively; cf., e. g., [12], p. 267. (The continuous spectrum $\Gamma(A)$ is the set of λ for which $c(\lambda) \geqq 1$.)

Let T be a closed symmetric operator with deficiency indices (m, m), where $0 < m \leqq \infty$. For a real λ, let

(11) $\pi(\lambda) = \min_A p(\lambda, A)$; (12) $\gamma(\lambda) = \min_A c(\lambda, A)$,

where the minimum is taken over all self-adjoint extensions A of T $(0 \leqq \pi(\lambda) \leqq \infty,\ 0 \leqq \gamma(\lambda) \leqq \infty)$.

(I) *If A is a self-adjoint extension of T, then*

(13) $\pi(\lambda) \leqq p(\lambda, A) \leqq \pi(\lambda) + m$; (14) $\gamma(\lambda) \leqq c(\lambda, A) \leqq \gamma(\lambda) + m$.

Proof. The relation (13) is a consequence of (β) and its proof; in fact, $\pi(\lambda)$ is the number π occurring in that assertion.

In order to prove (14), let A_1 be a self-adjoint extension of T such that

$$c(\lambda, A_1) = \gamma(\lambda). \tag{15}$$

It will be shown that, for every A,

$$c(\lambda, A_1) \geqq c(\lambda; A) - m. \tag{16}$$

It can be supposed that $m < \infty$ and that $c(\lambda, A) \geqq m$; for otherwise (16), which is equivalent to (14), is trivial.

Let $E(\lambda)$, $E_1(\lambda)$ denote the spectral resolutions of A, A_1, respectively. Let $y^1, y^2, \cdots$ denote (the possibly empty) set of eigenfunctions of A and let $\mathfrak{N}$ denote the closed manifold spanned by them. Let $x_1, x_2, \cdots$ be a sequence of elements with the properties that the sets $E(\lambda)x_k$, where $-\infty < \lambda < \infty$ and $k = 1, 2, \cdots$ span the closed linear manifold orthogonal to

$\mathfrak{N}$, that $(E(\lambda)x_k, x_j) = 0$ for $-\infty < \lambda < \infty$ and $j \neq k$, that $\sigma_k(\lambda) = \| E(\lambda)x_k \|^2$ is absolutely continuous with respect to $\sigma_{k-1}(\lambda)$. Then $c(\lambda, A)$ is the number of (continuous) functions $\sigma_1, \sigma_2, \cdots$ which are not constant on any open interval containing λ.

Let $\mathfrak{M}(x)$ denote the closed linear manifold spanned by the elements $E(\lambda)x$, where $-\infty < \lambda < \infty$. If $y_1, y_2, \cdots, y_m$ are m given elements orthogonal to $\mathfrak{N}$, then it can be supposed that the manifold spanned by $\mathfrak{M}(x_1), \cdots, \mathfrak{M}(x_m)$ contains $\mathfrak{M}(y_1), \cdots, \mathfrak{M}(y_m)$; cf. [12], pp. 247-262.

Let $\mathfrak{d}(l) \neq 0$ and let $x^1, x^2, \cdots, x^m$ be m linearly independent elements in $\mathfrak{M}(\bar{l})$. Let the elements $y_1, y_2, \cdots, y_m$ of the last paragraph be chosen to be the respective components of $x^1, \cdots, x^m$ in the manifold orthogonal to $\mathfrak{N}$.

If $k > m$, then $E(\lambda)x_k$ for every λ is orthogonal to $x^1, \cdots, x^m$; hence, the same is true of $(A - lI)^{-n}x_k$ for $n = 0, 1, \cdots$. By (δ), it follows that $(A_1 - lI)^{-n}x_k = (A - lI)^{-n}x_k$ for $n = 0, 1, \cdots$ and so, $E_1(\lambda)x_k = E(\lambda)x_k$ for $-\infty < \lambda < \infty$ and $k > m$. It follows from a standard procedure for determining $c(\lambda, A)$, that (16) and, therefore, (14) holds; cf. [12], pp. 247-262.

Remark. Let $\pi_0(\lambda) = \max p(\lambda, A)$ and $\gamma_0(\lambda) = \max c(\lambda, A)$, where the maxima are taken over all self-adjoint extensions A of T. Then $\pi(\lambda)$ and $\pi_0(\lambda)$ are, respectively, the dimensions of the manifolds determined by (7) and (3); that is, the multiplicities of λ in $\Pi(T)$ and $\Pi(T^*)$. For any integer k satisfying $\pi(\lambda) \leqq k \leqq \pi_0(\lambda)$, there is an $A = A(k, \lambda)$ such that $p(\lambda, A) = k$; cf. (α) and (β).

There naturally arises the question of the determination of $\gamma(\lambda)$ and $\gamma_0(\lambda)$, say in terms of T and T^*, and whether there exists an $A = A(k)$ such that $c(\lambda, A) = k$ when $\gamma(\lambda) \leqq k \leqq \gamma_0(\lambda)$. The answers to these questions would, in turn, answer the question raised by Weyl [14], pp. 251-252, in connection with his differential operators, as to whether or not the continuous spectrum $\Gamma(A)$ is independent of the boundary condition determining the self-adjoint extension A. They would also answer the somewhat more general question raised by Wintner as to the dependency of the continuous spectra $\Gamma(A + \tau E)$ on τ, where A is an arbitrary self-adjoint operator, E is a 1-dimensional projection and τ is a real number.

In this regard, (ii)-(vi) suggest that if $m < \infty$, then $\lambda = \mu$ is in every $\Gamma(A)$ if and only if $\lambda = 0$ is in $\Gamma((T^* - \mu I)(T - \mu T))$ and that $\gamma(\lambda) = \gamma_0(\lambda)$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[2] ——— and A. Wintner, "On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

[3] ——— and A. Wintner, "A separation theorem for continuous spectra," *ibid.*, vol. 71 (1949), pp. 650-662.

[4] ——— and A. Wintner, "Separation theorems for bounded Hermitian forms," *ibid.*, vol. 71 (1949), pp. 865-878.

[5] ——— and A. Wintner, "On the essential spectra of singular boundary value problems," *ibid.*, vol. 72 (1950), pp. 545-552.

[6] ——— and A. Wintner, "Lamé coordinates in Hilbert space," *ibid.*, vol. 72 (1950), pp. 775-791.

[7] E. Heinz, "Zur Theorie der Hermiteschen Operatoren des Hilbertschen Raumes," *Göttinger Nachrichten*, Math.-Phys. Kl. IIa, 1951, no. 2.

[8] E. Hellinger, "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen," *Journal für Mathematik*, vol. 136 (1909), pp. 210-271.

[9] B. von Sz. Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, Berlin (1942).

[10] J. von Neumann, "Ueber adjungierte Funktionaloperatoren," *Annals of Mathematics*, vol. 33 (1932), pp. 294-310.

[11] C. R. Putnam and A. Wintner, "The orthogonal group in Hilbert space," *American Journal of Mathematics*, vol. 74 (1952), pp. 52-78.

[12] M. H. Stone, *Linear transformations in Hilbert space and their applications to analysis*, New York (1932).

[13] H. Weyl, "Ueber beschränkte quadratische Formen, deren Differenz vollstetig ist," *Rendiconti del Circolo Matematico di Palermo*, vol. 27 (1909), pp. 373-392.

[14] ———, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[15] A. Wintner, "Zur Theorie der beschränkten Bilinearformen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 221-282.

[16] ———, *Spektraltheorie der unendlichen Matrizen*, Leipzig (1929).

[17] ———, "On non-singular bounded matrices," *American Journal of Mathematics*, vol. 54 (1932), pp. 145-149.

ON THE INFINITESIMAL GEOMETRY OF CURVES.*

By Aurel Wintner.

1. Let Γ be an arc of class C'' in the X-space, where $X=(x,y,z)$, that is, let Γ be a rectifiable Jordan arc having the property that the second derivative X'' of $X=X(s)$, where s is the arc length ($|X'|=1$), exists and is continuous. In particular, $\kappa=|X''|$ defines a continuous non-negative function $\kappa(s)$, the curvature on Γ. Suppose that $X''\neq 0$ on Γ; so that

$$\kappa=|X''|>0, \tag{1}$$

and so Frenet's mutually perpendicular unit vectors

$$U_1=X', \qquad U_2=\kappa^{-1}X'', \qquad U_3=[U_1,U_2] \tag{2}$$

(where the bracket refers to vector multiplication, hence $\det(U_1,U_2,U_3)$ is $+1$) exist and represent *continuous* functions $U_k(s)$ of s (in addition, $U_1(s)$ is of class C'). But Frenet's differential equations for (2) do not exist under the present assumption, since no torsion can be defined for an arbitrary Γ of class C'' satisfying (1).

The classical theory therefore assumes that Γ is of class C''' (i. e., that there exists a continuous third derivative $X'''(s)$). Then the torsion can be defined as $\det(X',X'',X''')/|X''|^2$ and is a continuous function, whereas the curvature $|X''|>0$ (instead of being just continuous as above) is a function of class C'. But the restriction of the theory of torsion to curves Γ of class C''' has disagreeable consequences in the theory of surfaces, for instance, if Γ is an asymptotic line or a geodesic, since, as pointed out in [2], p. 772 and [4], p. 608, respectively, the surface must then be required to be unnaturally smooth before the classical facts concerning these two types of curves on a surface can be formulated at all. In addition, one would surely like to (but, on the classical basis, cannot) say that a plane curve Γ has a torsion ($\equiv 0$) even if Γ is of class C'' only.

For these reasons, the following definition was introduced in [2], p. 771 (and will always be used in what follows): A curve Γ, of class C'' and of

* Received October 18, 1952; revised December 29, 1952.

non-vanishing curvature (1), is said to have a continuous torsion $\tau = \tau(s)$ if the vector functions (2) of s are such that

$$\tau = \lim_{\Delta s \to 0} \det(U_1, U_2, \Delta U_2/\Delta s) \tag{3}$$

where $U_k = U_k(s)$ and $\Delta U_2 = U_2(s + \Delta s) - U_2(s)$, exists and is a continuous function of s. The results of [2], pp. 773-774, imply that this will be the case if and only if

$$\lim_{\Delta s \to 0} \Delta U_2/\Delta s \tag{3 bis}$$

exists and is continuous, that is, if and only if $U_3 = [U_1, U_2]$ is a function of class C'. This means that (3) can be simplified to

$$\tau = \det(U_1, U_2, U_2'). \tag{4}$$

It may be mentioned that the above definition of *torsion* is the more natural from the geometrical point of view as it is the precise analogue of the definition of the Gaussian curvature, at non-parabolic points, of a surface of class C'' in terms of oriented spherical images. In fact, if a curve of class C'' has a non-vanishing curvature (and the latter itself is defined in terms of spherical images, those belonging to the tangent vector U_1), then (2) defines three continuous unit vectors. If the third of them, the binormal, is a function of class C' and s is not a stationary point, so that $U_3'(s) \neq 0$, then, near that point, $N = U_3(s)$ is a Jordan arc of class C' on the unit sphere $|N| = 1$, and the ratio of the oriented arc lengths of the portions $(s, s + \Delta s)$ of this curve and of the given curve, being equal to

$$\pm \int_s^{s+\Delta s} |U_3'(t)|\, dt / |\Delta s|,$$

tends to $\tau(s)$ as $\Delta s \to 0$. But nothing in this geometrical definition requires the usual assumption of a third derivative for $X(s)$.

2. We did not emphasize in [2] the following explicit criterion: *A curve of class C'' and of non-vanishing curvature possesses a continuous torsion if and only if all three unit vectors $U_k = U_k(s)$ are functions of class C'.* In fact, U_1 is of class C' whenever Γ is of class C'', and so the C'-character of U_3 follows from that of U_2, since $U_3 = [U_1, U_2]$, by (2).

Note that, in contrast to the situation in the C'''-theory, the curvature (1) and the torsion (4) are now of equal smoothness, namely, just continuous. As shown in [2], p. 772, the two continuous scalar functions $\kappa > 0$, τ and

the three continuously differentiable vector functions U_k satisfy Frenet's equations

$$(5)\qquad U_1' = \kappa U_2, \qquad U_2' = -\kappa U_1 + \tau U_3, \qquad U_3' = -\tau U_2$$

and, conversely, the assignment of any pair of continuous functions $\kappa > 0$, τ of s determines, via (5), a unique curve Γ: $X(s)$ of class C'' satisfying (1), (2) and (4) (the uniqueness of this $\Gamma = \Gamma(\kappa, \tau)$ is meant modulo the group of movements in the Euclidean X-space).

It was shown in [2] and [4], respectively, that this definition of a continuous torsion eliminates the above-mentioned objections to the classical theory of asymptotic curves and of geodesics. Furthermore, since $U_3(s) = \text{const.}$ in case of a plane curve of class C'' satisfying (2), it is clear from (4) that every such curve has the torsion $\tau(s) \equiv 0$ (the converse, too, is true; it follows from the existence and uniqueness theorems, mentioned after (5) above).

In terms of the approximate expansion of Γ near a point s of Γ, the replacement of the classical C'''-assumption by the more inclusive C''-class of continuous torsion can be characterized as follows: With reference to a fixed s, let the coordinate system $X = (x, y, z)$ be so chosen that $X(s) = (0, 0, 0)$ and that the coordinate axes x, y, z are those determined by the respective unit vectors $U_1(s)$, $U_2(s)$, $U_3(s)$. Then, under the assumption that Γ is of class C''' and of non-vanishing curvature, the approximate expressions of the three components of $X(s+h) - X(s) = X(s+h)$ are known to be

$$(5\text{ bis})\qquad h - \kappa^2 h^3/6, \qquad \tfrac{1}{2}\kappa h^2 + \kappa' h^3/6, \qquad \tau\kappa h^3/6,$$

respectively, with error terms *all three* of which are $o(|h|^3)$ as $h \to 0$ (the values of the functions κ, τ and of the continuous derivative κ' refer here to the point s). If the more inclusive case of Section 1 is considered, that is *if Γ is a curve of class C'' having a non-vanishing curvature and a continuous torsion, then, while the approximation* (5 bis), *with three $o(|h|^3)$-terms, need not hold, it holds in the curtailed form*

$$(5^*)\qquad h - \kappa^2 h^3/6 + o(|h|^3), \qquad \tfrac{1}{2}\kappa h^2 + o(h^2), \qquad \tau\kappa h^3/6 + o(|h|^3).$$

These asymptotic formulae, which can readily be deduced from (5), (2) and (1), suffice for the characterization of the values of $\kappa = \kappa(s) \neq 0$ and $\tau = \tau(s)$.

3. The precise nature of the classical C'''-assumption will now be analyzed explicitly, by proving the following criterion: *In order that a curve Γ of class C'' satisfying* (1) *be of class C''', the following pair of independent*

assumptions is necessary and sufficient: (i) *Γ has a continuous torsion $\tau=\tau(s)$ and* (ii) *the curvature $\kappa=\kappa(s)$ of Γ is of class C'.*

The necessity of both (i) and (ii) is the statement of the classical theory. The insufficiency of (i) alone is shown by the example of plane curves Γ which are of class C'' but not of class C'''. The insufficiency of (ii) alone can be concluded from the following example: On the closed interval $0 \leqq s \leqq 1$, let $\kappa(s)$ be a positive function possessing a continuous first derivative (e. g., $\kappa(s) \equiv 1$), and let $\tau(s)$ be a function which is bounded for $0 \leqq s \leqq 1$, continuous for $0 < s \leqq 1$ but discontinuous at $s=0$ (such as $\tau(s)=\sin 1/s$, where $s \neq 0$, and, for instance, $\tau(0)=0$). With reference to this pair, the linear differential equations (5) define on the interval $0 < s \leqq 1$ a curve $X=X(s)$ of class C'' satisfying (1), (2) and (4), where $U_1=X'$. Moreover, if (5_k) denotes the k-th of the three equations (5), application of two quadratures to (5_1) on the interval $\epsilon \leqq s \leqq 1$, when followed by the limit process $\epsilon \rightarrow 0$, shows that the limit $X(+0)$ exists and, if it is declared to be $X(0)$, then $X(s)$ is of class C'' on the closed interval $0 \leqq s \leqq 1$; and that (5_1), where $U_1=X'$ and $U_2=\kappa^{-1}X''$, holds at $s=0$ also. In particular, U_1 is of class C', and U_2 is continuous, for $0 \leqq s \leqq 1$. But this curve Γ: $X(s)$ cannot be of class C''' on $0 \leqq s \leqq 1$. For if it were, it ought to possess a continuous torsion at $s=0$ also. In particular, (5_3) would hold at $s=0$ with a continuous U_3'. Since (5_3) implies that $\tau=-U_2 \cdot U_3'$, this leads to the existence of the limit $\tau(+0)$, which contradicts the assumption.

This proves that (ii) alone is insufficient, and so it only remains to be shown that (ii) and (i) together are sufficient, in order that Γ be of class C'''. But this can be seen as follows: As pointed out above, (i) means that all three functions U_k of s are of class C'. In particular, U_2 is of class C'. It follows therefore from (5_1) and (ii) that U_1' is of class C'. Since $U_1'=X''$, this proves that X is of class C'''.

4. Let Γ: $X=X(s)$ be an arc of class C''' satisfying both (1) and

$$\tau \neq 0. \tag{6}$$

Then a classical result (cf. [1], pp. 28-40) states that, corresponding to every point s of Γ, there exist an unique "osculating sphere," that is, one and only one sphere having the property that the distance between the point $X(s+\Delta s)$ of Γ and that sphere is $o(|\Delta s|^3)$ as $\Delta s \rightarrow 0$. The vector, say Y, representing the center of this sphere is known to be

$$Y=X+\kappa^{-1}U_2-(\kappa^2\tau)^{-1}\kappa' U_3 \tag{7}$$

(cf. ibid.). Clearly, $Y = Y(s)$ is a continuous function. Let Γ_0 be the set of points described by $Y = Y(s)$ when s ranges over Γ. In what follows, it will be tacitly assumed that the locus $\Gamma_0 = \Gamma_0(\Gamma)$ is referred to a sufficiently short Γ. Although $Y(s)$ is a continuous function, Γ_0: $Y = Y(s)$ need not be a Jordan arc (for instance, Γ_0 is a single point when Γ is a circular arc).

In order to obtain an explicit condition preventing this contingency, suppose that Γ is of class C^4. Then $\kappa(s)$ will have a continuous second, and $\tau(s)$ a continuous first, and so the function (7) a continuous first, derivative. In order to obtain the latter, note that the derivative of the first term of (7) is $X' = U_1$, and the derivatives of the factors U_2, U_3 occurring in the second and third terms of (7) are given by (5_2) and (5_3). After trivial reductions, this supplies for the derivative $Y' = Y'(s)$ of $Y = Y(s)$ the representation

$$Y' = \sigma U_3, \tag{8}$$

if the scalar σ denotes the difference

$$\sigma = \tau\kappa^{-1} - (\tau^{-1}\kappa^{-2}\kappa')' \tag{9}$$

which, $X(s)$ being a function of class C^4 satisfying (1) and (6), is a continuous function $\sigma = \sigma(s)$.

If Γ: $X = X(s)$ satisfies the additional condition

$$\sigma \neq 0 \tag{10}$$

at a point s (hence near that point), then (8) shows that $Y' \neq 0$, which assures that Γ_0: $Y = Y(s)$ is a Jordan arc and, what is more, that Γ_0 is an arc possessing a continuous tangent (in fact, Y' is a continuous function of s).

A Jordan arc Γ: $X = X(s)$ will be called *a curve of class C^n free of spherical points* [1] if the function $X(s)$ has a continuous n-th derivative and satisfies (10) for every s. This implies that $n \geqq 4$ and that neither the curvature κ nor the torsion τ vanishes on Γ, since otherwise the function (9) cannot even be defined.

---

[1] In order to arrive at an interpretation of assumption (10), or rather of its violation, $\sigma(s) = 0$, at a *single* point s, it is sufficient to observe that $\sigma(s) \equiv 0$, the *identical* vanishing of the function $\sigma(s)$, is characteristic of those curves Γ of class C^4 satisfying (1) and (6) which are situated on a fixed sphere (cf. [1], p. 41). A point s of Γ can therefore be called a "spherical" or a "non-spherical" point according as it violates or satisfies condition (10). The situation becomes clear if, corresponding to this terminology, a point s of Γ is called a "planar" or a "non-planar" point according as it violates or satisfies condition (6) (where only (1) and the C'''-character of Γ need be assumed).

5. If the torsion of a curve is defined in the classical manner (that is to say so as to assume the C'''-character of the curve), then, in order to be able to speak of the torsion of the curve Γ_0, the curve Γ is restricted to be of class C^6, since 3 degrees of differentiability appear to be sacrificed in the passage from $X(s)$ to $Y(s)$; cf. (7). But it turns out that, on the one hand, this appearance is misleading, since $n=6$ can be reduced to $n=4$, and that, on the other hand, definition (4) of the torsion allows a reduction by one more degree of differentiability. In other words, the restriction to the class C^6, which is a tacit assumption of Monge's theory (cf. [1], pp. 42-43), can actually be reduced to the class C^4 which, by the end of Section 4, represents the optimum in this context (at least if $\sigma=\sigma(s)$, the "measure of non-spherical character," is defined by the explicit formula (9)). In fact, it will be proved that *if a curve* Γ *of class* C^4 *is free of spherical points, then* Γ_0, *the locus of the centers of the osculating spheres of* Γ, *is a curve of class* C'' *possessing a non-vanishing continuous curvature and a non-vanishing continuous torsion.* If the latter are denoted by κ_0 and τ_0, respectively, their explicit representations in terms of curvature and the torsion of Γ itself prove to be, as in the classical case (cf. [1], pp. 42-43),

$$(11) \qquad \kappa_0=|\tau/\sigma|, \qquad\qquad (12) \qquad \tau_0=\kappa/\sigma.$$

6. The first assertion of the italicized theorem, namely, the C''-character of Γ_0, can be proved as follows: In view of (10), the quadrature assigned by

$$(13) \qquad ds_0=\sigma ds,$$

where $\sigma=\sigma(s)$ is continuous, establishes between the arc length s on Γ and the parameter s_0 a one-to-one correspondence in such a way that the function $s_0=s_0(s)$ and its inverse $s=s(s_0)$ have continuous first derivatives ($\neq 0$) with respect to s and s_0, respectively. It is also seen from (13) and (10) that, generally,

$$(14) \qquad F'=\sigma F^{\backprime}, \text{ where } \sigma\neq 0 \text{ and } F^{\backprime}=dF/ds_0, \qquad F'=dF/ds.$$

Hence, (8) can be written in the form

$$(15) \qquad Y^{\backprime}=U_3.$$

In view of $|U_3|=1$, this implies that $|Y^{\backprime}|=1$, which means that s_0 is the arc length on Γ_0. It also follows from (15) that, in order to prove that $Y=Y(s_0)$ is of class C'', it is sufficient to ascertain that U_3 is of class C' as a function of s_0. But this is obvious, since $U_3=U_3(s)$ and $s=s(s_0)$

are functions of class C' of s and s_0, respectively. This proves that Γ_0 is an arc of class C''.

Since (15) and (14) imply that $Y^{``} = \sigma^{-1} U_3'$, it follows from ($5_3$) that

$$(16) \qquad Y^{``} = -\sigma^{-1}\tau U_2,$$

hence $|Y''| = |\tau/\sigma|$. This proves (11). But (11) implies, by (6), that $\kappa_0 \neq 0$. Consequently, there belongs to Γ_0: $Y = Y(s_0)$ three unit vectors, say $V_k = V_k(s_0)$, in the same way as the three unit vectors (2) belong to Γ: $X = X(s)$; so that

$$(17) \qquad V_1 = Y^{`}, \qquad V_2 = \kappa_0^{-1} Y^{``}, \qquad V_3 = [V_1, V_2].$$

But (16) and (11) show that $Y^{``} = \pm \kappa_0 U_2$, where

$$(18) \qquad \pm = \operatorname{sgn}(-\tau/\sigma).$$

If this and (15), respectively, are substituted into the second and the first of the relations (17), it follows that

$$(19) \qquad V_1 = U_3, \qquad V_2 = \pm U_2, \qquad -V_3 = \pm U_1,$$

since the matrices $\| U_1, U_2, U_3 \|$, $\| V_1, V_2, V_3 \|$ are orthogonal and of determinant $+1$.

Since every U_k is of class C' (as a matter of fact, of class $C^{4-2} = C''$) as a function of s, and since $s = s(s_0)$ is a function of class C' in s_0, it follows from (19) that all three functions $V_k = V_k(s_0)$ have continuous first derivatives $V_k^{`}$. In view of the criterion italicized at the beginning of Section 2, this proves that Γ_0 has a continuous torsion $\tau_0 = \tau_0(s_0)$.

Accordingly, only the explicit formula (12) remains to be proved. To this end, it is sufficient to apply (4) to the present case. In fact, this gives $\tau_0 = \det(V_1, V_2, V_2^{`})$, hence $\tau_0 = \det(U_3, U_2, U_2^{`})$, where $U_2^{`} = \sigma^{-1} U_2'$, by (14). It follows therefore from (5_2) that $\tau_0 = \det(U_3, U_2, -\sigma^{-1}\kappa U_1)$. In view of (4), this proves (12).

7. In view of the comment at the end of Section 4, it is worth pointing out that in the case

$$(20) \qquad \kappa(s) = \mathrm{k}, \quad \text{where } \mathrm{k} = \text{const.} > 0,$$

a case specifically discussed by Monge (cf. [1], pp. 43-44), the C^4-assumption of the last italicized theorem case be reduced to a C^3-assumption, and even to a C^2-assumption with a continuous torsion: *Γ_0 is of class C'' possessing a continuous curvature and a continuous non-vanishing torsion whenever Γ is of*

class C'' and possesses a constant curvature and a continuous non-vanishing torsion. In addition, (19) holds again, and the curvature $\kappa_0 = \kappa_0(s)$ and the torsion $\tau_0 = \tau_0(s)$ of Γ_0 are given, in terms of the curvature (20) and the torsion $\tau = \tau(s)$ of Γ, by

$$(21) \qquad \kappa_0(s) = \mathrm{k}, \qquad\qquad (22) \qquad \tau_0(s) = \mathrm{k}^2/\tau(s).$$

The assumption (20) reduces the definition (9) to

$$(23) \qquad \sigma = \tau/\mathrm{k},$$

which in turn reduces (11)-(12) to (21)-(22). But this formal conclusion is not legitimate, since (23) depends on (9), and therefore on the C^4-assumption of Section 5-6. But if the continuous function $\sigma = \sigma(s)$ is *defined* by (23) under the assumption (20), then the last italicized statement follows by a straightforward repetition of the proof given in Section 6.

8. Let Γ: $X = X(s)$ be an arc of class C'' satisfying (1). Then there is at every point s of Γ an unique osculating circle. The latter has the radius $1/\kappa(s)$, and its center, say $Z = Z(s)$, is situated on the positively oriented principal normal $U_2 = U_2(s)$; so that

$$(24) \qquad Z = X + \kappa^{-1} U_2.$$

Let Γ^* denote the locus of all points (24) when s varies on Γ. It will be assumed that Γ is sufficiently short.

If Γ is of class C''', then κ' exists and is continuous (cf. Section 3) and the vertices of Γ (if any) are defined as the points s at which κ becomes stationary, $\kappa' = 0$. Let it be assumed that either

$$(25) \qquad \kappa' \neq 0$$

or (6) holds, i. e., that τ and κ' do not vanish simultaneously. Then $Z' \neq 0$ holds at every point of Γ. In fact, if (24) is differentiated and X' and U_2' are then substituted from (2) and (5_2), it follows that

$$(26) \qquad Z' = -\kappa'\kappa^{-2} U_2 + \tau\kappa^{-1} U_3.$$

Suppose in particular that Γ is a plane curve. Then (6) is violated identically and (26) reduces to

$$(27) \qquad Z' = -\kappa'\kappa^{-2} U_2.$$

Since (2_2) and (1) show that the representation (24) of the evolute Γ^* of Γ is

$$(28) \qquad \Gamma^*\colon Z = X + |X''|^{-2} X'',$$

two degrees of differentiability appear to be lost in the passage from Γ to Γ^*; so that Γ^* seems to be of class C' only, if Γ is just of class C'''. But it turns out that *one* of the two degrees of differentiability is *not* lost under the present assumptions, i. e., that the situation is as follows: *If a plane curve* Γ *of class* C''' *has a non-vanishing curvature and is free of vertices, then its evolute* Γ^* *is a curve of class* C''. In particular, Γ^* has a continuous curvature $\kappa^* = \kappa^*(s)$. The latter does not vanish, since its explicit representation proves to be the same as under the standard assumptions, namely, $\kappa^* = |\kappa'|/\kappa$.

All of this can be proved by adapting to (27) the procedures applied in Section 6. The proof will be omitted for this reason and also because the last italicized statement will follow in Sections 10-12 as a corollary of a more general theorem.

The cusps of the evolute of an ellipse show that the curve Γ^* need not be of class C' (and still less, as claimed by the assertion, of class C'') if the restriction (25) is omitted (if the ellipse is a circle, $\kappa' \equiv 0$, then Γ^* is not even a curve).

Actually, an easy calculation shows that if (25) is omitted from the assumptions of the last italicized theorem, and if P^* is the point of Γ^* corresponding to a point P of Γ, then Γ^* *must* have a cusp at P^* if κ' changes sign at P and does not vanish at points of Γ close enough to, but distinct from, P.

It should be noted that, in the last italicized statement, restriction of Γ to plane curves is essential, i. e., that *the curve* Γ^*: $Z = Z(s)$, defined by (24), need not be of class C'' if Γ: $X = X(s)$ is a twisted curve of class C''', which has a non-vanishing curvature and is free of vertices. It is understood that a vertex can be defined as a point s of Γ at which the derivative $Z'(s)$ of (24) vanishes, which, in view of (26), requires that neither of the conditions (6), (25) be satisfied. Actually, a counterexample can be chosen so as to satisfy both (6) and (25).

9. Let S: $X = X(u, v)$ be a surface of class C'''. Then the Gaussian and mean curvatures are functions, $K = K(u, v)$ and $H = H(u, v)$, of class C' satisfying the inequality $H^2 \geqq K$. Let S be free of umbilical points, that is, let

$$(29) \qquad H^2 > K.$$

Then the principal curvatures k_1, k_2, being the roots of the quadratic equation $k^2 - 2Hk + K = 0$, are distinct and of class C'. Hence, at least one of these curvatures does not vanish (if (u, v) is confined to a sufficiently small

vicinity of any fixed (u^0, v^0)). The corresponding principal radius of curvature $\rho = 1/k$ is a function $\rho(u, v)$ of class C'.

The lines of curvature on S are defined, after a suitable rotation of the (u, v)-plane, by two differential equations of the form

$$dv/du = f(u, v), \tag{30}$$

where f is of class C', and consist of two transversal families, each of class C' and each covering S in a *schlicht* manner, if S is sufficiently small. By virtue of the equation of Rodrigues,

$$kdX + dN = 0, \tag{31}$$

there belongs to each principal curvature k *one* of these two families of curves, in the sense that (31) will hold along the curves of the family if sgn H, hence sgn k, and the orientation of the unit normal $N = N(u, v)$ are suitably chosen. If $k \neq 0$ and $\rho = 1/k$, then (31) can be written as

$$dX + \rho dN = 0. \tag{32}$$

While (32) is valid for at least one family of lines of curvature, that belonging to a non-vanishing principal curvature, it might not be possible to write the equation (31) corresponding to the other principal curvature in the form (32), since the vanishing (or even the identical vanishing) of the latter curvature is allowed, that is, the Gaussian curvature $K = k_1 k_2$ can satisfy $K(u, v) = 0$ (or even $K(u, v) \equiv 0$).

It will be assumed that sgn $\rho (\neq 0)$ and the orientation of N have been chosen so that (32) holds; in particular, the orientation of ρN is uniquely determined. The vector $Y = Y(u, v)$ defined by

$$Y(u, v) = X(u, v) + \rho(u, v) N(u, v) \tag{33}$$

is of class C'. The locus T: $Y = Y(u, v)$ need not be a surface of class C', since the condition

$$[Y_u, Y_v] \neq 0 \tag{34}$$

can be violated. However, it will be shown that if this condition is not violated, that is, *if* (33) *is a C'-parametrization of a surface T, then T must be a surface of class C''.* This does not mean, of course, that the *function* (33) is of class C'', but merely that the surface T, which is one of the two classical evolutes of the surface S (the evolute belonging to that root $\rho = \rho(u, v)$ of the quadratic [possibly linear] equation $K\rho^2 - 2H\rho + 1 = 0$ which occurs in the definition (33) of T) admits of *some* parametrization, say $T: Y = Y(u^*; v^*)$, in which it becomes of class C''.

Although (33) is a function of class C', the evolute T represented by it need not be a surface of class C', even if distinct points (u, v) correspond to distinct points $Y(u, v)$. This is shown by the example of a cylinder having an elliptical (not a circular) cross-section; cf. Section 10 below.

In order to assure that T is a surface of class C', it is sufficient to assume (29) and

$$\rho_\nu \neq 0, \tag{35}$$

where $\rho_\nu = \partial\rho/\partial\nu$ denotes the derivative of $\rho = \rho(u, v)$ in the direction of that line of curvature, passing through the point (u, v) of S, which is in the family of lines of curvature belonging to ρ by virtue of the normalization (32). In fact, (34) holds if and only if (29) and (35) hold (cf. [1], pp. 417-418). Accordingly, the last italicized statement is equivalent to the case $n = 3$ of the following theorem:

On a surface S: $X = X(u, v)$ of class C^n, where $n \geqq 3$, let (29) *hold and let $\rho = \rho(u, v)$ be a (finite) radius of principal curvature satisfying* (35). *Then the corresponding evolute T: $Y = Y(u, v)$, defined by* (35), *is a surface of class C^{n-1} (locally).*

This theorem is an analogue of theorem (iii) in [5], p. 368. The latter theorem replaces the evolute surface (33) by a parallel surface $Z = Z(u, v) = Z(u, v; r)$,

$$Z = X(u, v) + rN(u, v), \quad \text{where} \quad r = \text{const.} \gtrless 0, \tag{33*}$$

and claims that, except when the constant r satisfies the quadratic (possibly linear) equation $K(u, v)r^2 - 2H(u, v)r + 1 = 0$ at some point (u, v) of the (small) surface S, the parallel *surface* P_r: $Z = Z(u, v; r)$ is "smoother" than is indicated by the defining equation (33*). Curiously enough, the restriction placed on the value of the constant r, a restriction rendering P_r a surface of class C' (and playing, therefore, the same rôle as do the inequalities (29) and (35) above), is precisely $\rho(u, v) \neq r$, where $\rho(u, v)$ is a (finite) principal curvature, as in (33).

10. The content and the nature of the assertions of the last theorem can well be illustrated by the following corollary of it:

In an X-plane, where $X = (x, y)$, let S: $X = X(s)$ be an arc of class C^n, where s is arc length and $n \geqq 3$, and suppose that the curvature $\kappa = |X''|$ ($= |N'|$, where N denotes $(-y', x')$, the normal vector) does not vanish and does not become stationary; that is,

(29 bis) $$\kappa(s) > 0$$

and

(35 bis) $$\kappa'(s) \neq 0.$$

Then the evolute T: $Y = Y(s)$, *defined by*

(33 bis) $$Y(s) = X(s) + \rho(s)N(s), \quad \text{where} \quad \rho = 1/\kappa,$$

is a curve of class C^{n-1} (*locally*), even though the function $Y(s)$ is of class C^{n-2} only (unless $X(s)$ is of class C^{n+1}).

In fact, if V denotes the unit vector $(0, 0, 1)$ and if the plane vector $X = (x, y)$ is thought of as the vector $(x, y, 0)$, then $X(s, t) = X(s) + tV$ is a cylinder of class C^n. Condition (29 bis) assures that (29) holds on this surface; in fact, the principal curvatures k_1, k_2 on this cylinder are $k = \kappa > 0$ and $k_2 \equiv 0$. The family of lines curvature belonging to $\rho = 1/k_1 = 1/\kappa$ are the curves $t = \text{const.}$, that is, $X = X(s)$ and its congruent images $X = X(s) + tV$. Hence, condition (35) is equivalent to (35 bis). The evolute (33) belonging to the principal radius of curvature $\rho = 1/\kappa$ is the cylinder having the evolute (33 bis) as it cross-section; that is, the cylinder

$$T\colon\ Y(s,t) = Y(s) + tV \equiv X(s) + \rho(s)N(s) + tV.$$

11. In the following proof of the theorem italicized in Section 9, it will be assumed that $n = 3$. The proof will be such as to make clear its validity for $n > 3$ also.

Since S: $X = X(u, v)$ is of class C''' and (29) holds, a sufficiently small neighborhood of a point (u^0, v^0) can be transformed by a mapping $(u, v) \to (\alpha, \beta)$ of class C' and of non-vanishing Jacobian in such a way that the two families of the lines of curvature on S become represented by $\alpha = \text{const.}$ and $\beta = \text{Const.}$; cf. the above remarks concerning (30). Let $X = X(\alpha; \beta)$ denote the function which results if the functions $u = u(\alpha, \beta)$, $v = v(\alpha, \beta)$ of class C' are substituted into the function $X(u, v)$ (of class C'''). Thus $X = X(\alpha; \beta)$ is a C'-parametrization of the surface S (of class C'''), with the lines of curvature of S as parameter lines. Note that (as shown in [3], pp. 168-172) such a parametrization cannot, in general, be of class C'''; probably, it need not be even of class C''.

The normal $N(\alpha; \beta) = [X_\alpha, X_\beta]/|[X_\alpha, X_\beta]|$, which apparently is just continuous, is actually of class C' (as a function of (α, β)). For, on the one hand, $N(\alpha; \beta) \equiv \pm N(u, v)$ by virtue of $u = u(\alpha, \beta)$, $v = v(\alpha, \beta)$, while, on the other hand, $N(u, v)$ is a function of class C' (as a matter of fact, C'') as a

function of (u, v), and $u = u(\alpha, \beta)$, $v = v(\alpha, \beta)$ are functions of class C'. The principal curvatures $k_1(\alpha; \beta)$, $k_2(\alpha; \beta)$ and the principal radius of curvature $\rho(\alpha; \beta)$ are defined by invariance (for example, $\rho(\alpha; \beta) \equiv \pm \rho(u, v)$) and are therefore of class C'. The orientation of ρN is chosen so as to satisfy (32).

Although the (non-vanishing) perpendicular vectors X_α, X_β might only be continuous functions, the corresponding unit vectors $X_\alpha / |X_\alpha|$, $X_\beta / |X_\beta|$ are of class C' as functions of (α, β). In order to see this, note that $X_\alpha = X_u u_\alpha + X_v v_\alpha$. If (30) is the differential equation defining the lines of curvature $\beta = \text{const.}$, then $u_\alpha \neq 0$, and X_α is parallel to the non-vanishing vector $X_u + X_v(v_\alpha / u_\alpha) = X_u + X_v f(u, v)$. The latter is of class C' as a function of (u, v), and so the same is true of it as a function of (α, β). Consequently, $X_\alpha / |X_\alpha|$ (and similarly $X_\beta / |X_\beta|$) is of class C'.

Let α, β, $X(\alpha; \beta)$, $N(\alpha; \beta)$ be renamed u, v, $X(u, v)$, $N(u, v)$, respectively. Then $X = X(u, v)$ is just a C'-parametrization of the surface S of class C''', but $N(u, v)$ is of class C', and $u = \text{const.}$, $v = \text{Const.}$ are the two families of lines of curvature. Let the notation be so chosen that $v = \text{Const.}$ is the family which, in the sense specified after (32), belongs to the root $\rho = \rho(u, v)$ occurring in (33). Then (32) shows that the differential equation defining the lines of curvature $v = \text{Const.}$ is

$$X_u + \rho N_u = 0, \tag{36}$$

where, according to (35),

$$\rho_u \neq 0. \tag{37}$$

It also follows from (31) that, if $k_1 = k_1(u, v)$ and $k_2 = k_2(u, v)$ are the principal radii of curvature on S, and if the notation is so chosen that $\rho = 1/k_1$ in (36), then the equation (31) of the family $u = \text{const.}$ is

$$k_2 X_v + N_v = 0. \tag{38}$$

Finally, as verified above,

$$X_u / |X_u| \text{ is of class } C'. \tag{39}$$

12. Since all three functions X, N, ρ of (u, v) are of class C', the same is true of the function (33). But differentiation of (33) shows that

$$\tilde{Y}_u = \rho_u N \text{ and } Y_v = (1 - \rho k_2) X_v + \rho_u N \tag{40}$$

by virtue of (36) and (38), respectively. Hence it is easy to see that (34) holds.

In fact, the vector product of the derivatives (40) is

$$[Y_u, Y_v] = \rho_u(1-\rho k_2)[N, X_v]. \tag{41}$$

The vector $[N, X_v]$ does not vanish; in fact, since $u = \text{const.}$ and $v = \text{Const.}$ are lines of curvature, the non-vanishing vectors X_u, X_v, N are mutually perpendicular. Hence,

$$[N, X_v] \neq 0 \text{ is parallel to } X_u/|X_u|. \tag{42}$$

The factor $1-\rho k_2$ in (41) is not 0, since the principal curvatures $k_1 = 1/\rho$, k_2 are distinct. Hence (34) follows from (41) and (37) and, since (41) and (42) imply that

$$[Y_u, Y_v]/|[Y_u, Y_v]| = \pm X_u/|X_u|, \tag{43}$$

T is a surface of class C'.

In addition, (39) shows that T is a surface possessing a C'-parametrization $Y = Y(u, v)$ in which the unit normal (43) is of class C'. Hence, in order to conclude that, as claimed by the italicized theorem of Section 9, the surface T is of class C'', it is sufficient to apply the argument used at the end of Section 14 in [3], p. 163.

13. After the preceding generalization of the italicized statement of Section 8, a question complementary to that treated in Sections 4-6 will now be considered. In fact, whereas (10) was there assumed for $X = X(s)$ at every s, the identical violation of (10) will now be dealt with.

Assumption (10) for the function $\sigma = \sigma(s)$ defined by (9) is the natural condition for the non-degeneracy of the locus of the osculating spheres of a curve Γ: $X = X(s)$ and, correspondingly, the *identical* violation of (10) is the classical condition for a *spherical* Γ, that is, for a Γ satisfying $|X(s)| = \text{const.}$; cf. the footnote in Section 4. But this classical characterization of a Γ situated on a sphere (that is, the differential equation

$$(\kappa^{-2}\kappa'\lambda)'\lambda = 1, \quad \text{where} \quad \lambda = \kappa/\tau, \tag{44}$$

which, in view of (9), is equivalent to $\sigma \equiv 0$) assumes, on the one hand, that κ and τ are of class C'' and C', respectively, implying for Γ an unnaturally strong C^4-restriction, and excludes, on the other hand, the cases of clustering or isolated zeros of the torsion (not to mention the case $\tau \equiv 0$ of a great circle Γ on the sphere $|X| = \text{const.}$); needless to say, $\tau(s) \not\equiv 0$ can have clustering zeros even if Γ is of class C^∞. In what follows, there will be derived a criterion which, though entirely explicit, is free of the artificial restrictions assumed in the classical criterion (44).

In order to simplify the formulae, suppose first that the radius of the sphere containing Γ: $X = X(s)$ is normalized to be 1,

$$(45)\qquad X^2 = 1.$$

If Γ is of class C'' (so that τ need not even exist), two differentiations of (45) give

$$(46)\qquad X \cdot U_1 = 0, \qquad X \cdot U_2 = -1/\kappa.$$

In fact, (2) is applicable under the C''-assumption alone, if the inequality (1) is satisfied. But it is, since the second derivative of (45) is $X'^2 + X \cdot X'' = 0$ which, in view of $|X'| = 1$, prevents the vanishing of X''. Actually, since X and U_2 are unit vectors, it follows from (46_2) that the inequality (1) can be improved to

$$(47)\qquad \kappa \geqq 1,$$

where the sign of equality holds if and only if $X \cdot U_2 = -1$ (which means that

$$(48)\qquad \kappa(s^0) = 1 \text{ if and only if } X(s^0) = -U_2(s^0)$$

at some $s = s^0$).

On the other hand, since X is a linear combination of the three vectors (2), it follows from (46_1) that there exist two (continuous) scalar functions $\alpha = \alpha(s)$, $\beta = \beta(s)$ satisfying

$$(49)\qquad X = \alpha U_2 + \beta U_3, \qquad \alpha^2 + \beta^2 = 1,$$

(49_2) being a consequence of (49_1) and (45) (since $|U_i| = 1$ and $U_2 \cdot U_3 = 0$). The coefficients of (49_1) are given by

$$(50)\qquad \alpha = -1/\kappa, \qquad \pm\beta = (1 - \kappa^{-2})^{\frac{1}{2}}; \quad \text{cf. } (47).$$

In fact, (49_1) and (46_2) imply (50_1), whence (50_2) follows by (49_2). The alternative sign in (50_2) remains undecided and can, by continuity, change only at points $s = s^0$ at which

$$(51)\qquad \beta(s^0) = 0, \quad \text{i.e.,} \quad \kappa(s^0) = 1 \text{ or } X(s^0) = -U_2(s^0);$$

cf. (50_2) and (48).

14. Needless to say, (49) and (50) are not only necessary but sufficient as well for a Γ of class C'' satisfying (45). It will now be assumed that such a Γ *satisfies the additional assumption of possessing a continuous torsion* $\tau = \tau(s)$ in the sense of Section 1. Under this assumption, it will be shown that

(I) *Γ must be of class C''' (i. e., κ' exists and is continuous)*;

(II) *$\kappa' = 0$ at all those points at which $\kappa = 1$*;

(III) *κ'' exists (and is non-negative) at all those points at which $\kappa = 1$*;

(IV) *the absolute value of $\tau = \tau(s)$ can be calculated from $\kappa = \kappa(s)$; in fact,*

$$\pm \tau = \kappa'/(\kappa^4 - \kappa^2)^{\frac{1}{2}} \text{ if } \kappa > 1, \tag{52}$$

$$\pm \tau = \kappa''^{\frac{1}{2}} \text{ if } \kappa = 1; \quad \text{cf. (III)}. \tag{53}$$

First, since Γ is supposed to have a continuous torsion, the functions U_i of s are of class C' (Section 2). Hence, scalar multiplication of (49_1) by U_2, U_3 proves the C'-character of α, β, respectively. It follows therefore from (50_1) that κ is a function of class C', and from (50_2), that the same is true of $\pm(\kappa^2 - 1)^{\frac{1}{2}}$ when the choice of the alternative sign is suitably made at points $s = s^0$ satisfying (48). The first of the latter two conclusions proves (I) if use is made of the italicized result of Section 3, while the second conclusion assures that, if $\kappa(s^0) = 1$ and $s^0 = 0$, then

$$\pm(\kappa^2 - 1)^{\frac{1}{2}} = cs + o(|s|) \text{ as } s \to 0 \tag{54}$$

holds for some constant c. The alternative sign in (54) is undecided and might change when s passes through 0. In any case $\kappa^2 - 1 = c^2s^2 + o(s^2)$, hence

$$\kappa = 1 + \tfrac{1}{2}c^2s^2 + o(s^2), \qquad \kappa' = o(1) \tag{55}$$

(in fact, (55_2) is implied by (55_1), since κ' exists and is continuous). Clearly, (55_1), (55_2) prove (III), (II), respectively.

Next, if the identity (49_1) is differentiated and X', U_2', U_3' are then substituted from (2_1), (5_2), (5_3), respectively, comparison of the coefficients of U_1, U_2, U_3 in the resulting identity supplies the three relations (50_1),

$$\alpha' - \beta\tau = 0, \qquad \beta' + \alpha\tau = 0. \tag{56}$$

But (50) and either (56_i) imply assertion (52) of (IV). Finally, assertion (53) of (IV) can be proved as follows:

According to (50_2), the function β is $1/\kappa$ times $\pm(\kappa^2 - 1)^{\frac{1}{2}}$. If this product is differentiated, it follows from (54) that

$$\beta' = -(cs + o(|s|))\kappa'/\kappa^2 + (c + o(1))/\kappa.$$

Since $\kappa \to 1$ as $s \to 0$, this implies that $\beta' \to c$, which, in view of (56_2), means that $\alpha\tau \to -c$. It follows therefore from (50_1) that $\tau = c$ at $s = 0$, which, in view of (55), proves (53).

15. It is now easy to verify the following improvement of the C^4-criterion (44) (in a way which, in contrast to (44), does not exclude the important possibility $\tau = 0$) : A Γ: $X = X(s)$ *of class* C'' *possessing a continuous torsion is a curve on the unit sphere if and only if it is of class* C''' *and its curvature and torsion satisfy* (47) *and* (52)-(53).

Needless to say, the normaliaztion of a spherical curve to be of radius 1 is unessential, since the theorem remains unaltered if τ and κ in (52)-(53) are replaced by $r\tau$ and $r\kappa$, respectively, when (45) is replaced by $X^2(s) \equiv r^2$, where r is any positive constant.

It is clear from (I)-(IV), Section 14, that the last italicized theorem will follow if it is ascertained that a Γ: $X = X(s)$ of class C''' must satisfy (45) whenever its curvature and torsion are subject to (47) and (52)-(53). But (5) and the assignment of any pair of continuous functions $\tau(s), \kappa(s) > 0$ determine for (5) a solution (U_1, U_2, U_3) of orthonormal vectors (of determinant $+1$) which is unique modulo the group of euclidean movements. On the other hand, if $\tau(s)$ is given, in terms of a $\kappa(s)$, by (52)-(53), then, by retracing the steps made in Sections 13-14, it can be verified that the $X = X(s)$ defined by (49) and (50) must satisfy (2) and (5). This proves the last italicized statement.

16. In the direction of the results of this paper, all of which deal with the possibility of savings in the assumed degree differentiability, a final remark will now be made on the Legendre transformation of a curve Γ: $y = y(x)$ in an (x, y)-plane, that is, *on the replacement of the point coordinates of a plane curve by its line coordinates.* Formally, this consists of the replacement of x, y by ξ, η, where, if $' = d/dx$,

$$(57_1) \qquad \xi = y', \qquad\qquad (57_2) \qquad \eta = xy' - y,$$

and of the assertion that the inverse of the substitution $(x, y) \rightarrow (\xi, \eta)$ is

$$(58_1) \qquad x = \eta^{\backprime}, \qquad\qquad (58_2) \qquad y = \xi\eta^{\backprime} - \eta,$$

where $^{\backprime} = d/d\xi$.

In order to formulate conditions under which this formalism is valid, it must first be assumed that $y(x)$ has on some x-interval (a, b) a derivative $y'(x)$ (in some sense) and that, if (57_1) maps (a, b) onto the ξ-interval (α, β), the functions (57_i), (58_i) are considered on (a, b), (α, β), respectively. Then the classical theorem on Legendre's transformation can be

formulated as follows (cf. [6], pp. 6-8, where $x, y; \xi, \eta$ are vectors with n components; so that the present case results by choosing $n = 1$ *loc. cit.*):

Suppose that $y(x)$ is of class C'' and that, in addition, its second derivative (the "Hessian" of $y(x)$, i. e., the "Jacobian" of (57_1)) is subject to the restriction

$$(59) \qquad y'' \neq 0$$

on (a, b). Then it is clear that (57_1) maps (a, b) onto (α, β) in a one-to-one C'-manner and in such a way that the inverse mapping $\xi \rightarrow x$, too, is of class C'. The classical theorem states that if this C'-function $x = x(\xi)$ is substituted into the C'-function (57_2) of x, the resulting function $\eta = \eta(\xi)$ will be of class C'' on (α, β), and that (58_1)-(58_2) is the explicit form of the inverse of the mapping (57_1)-(57_2) of (x, y) onto (ξ, η).

Notice however that the C''-assumption on $y(x)$ and the restriction (59) are only sufficient in order that the mapping (57_1) of (a, b) onto (α, β) be continuous and one-to-one, i. e., for the following condition:

$$(60) \qquad y'(x) \text{ is strictly monotone} \qquad ('=d/dx)$$

on (a, b). In fact, if

$$(61_1) \qquad y(x) = |\, x^3 \,|^{\frac{1}{3}}, \qquad (61_2) \qquad y(x) = x^4,$$

then both functions (61_i) are of class C' and such as to satisfy (60) but, if $a < 0 < b$, the function (61_1) fails to be of class C'' while the function (61_2) fails to satisfy (59). These examples reveal the content of the following extension (applicable to both (60_1) and (60_2)) of the classical theorem:

If $y(x)$ is a differentiable function satisfying (60) on (a, b) (which implies that $y(x)$ is of class C'), and if the inverse $x = x(\xi)$ of the function (57_1) of x is substituted into the function (57_2) of x, then the resulting function $\eta = \eta(\xi)$ is differentiable on the ξ-interval (α, β) which corresponds to (a, b), and

$$(60 \text{ bis}) \qquad \eta^{`}(\xi) \text{ is strictly monotone} \qquad (^{`} = d/d\xi)$$

on (α, β) (which implies that η is of class C' as a function of ξ); finally, the explicit form of the inverse of the contact transformation $(x, y) \rightarrow (\xi, \eta)$, defined by (57_1)-(57_2), is given by (58_1)-(58_2).

The proof of this somewhat unexpected extension of the classical theorem depends only on careful applications of the definition ($= \lim \Delta v / \Delta u$) of a derivative, and will be omitted. Actually, this theorem is just a degenerate case (dealing with plane curves) of Theorem (II) (dealing with surfaces) in a paper of Hartman and myself, to appear in this JOURNAL.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] L. Bianchi, *Lezioni di geometria differenziale*, 3rd ed., vol. I (1927).

[2] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[3] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[4] ——— and A. Wintner, "On geodesic torsions and parabolic and asymptotic arcs," *ibid.*, vol. 74 (1952), pp. 607-625.

[5] A. Wintner, "On parallel surfaces," *ibid.*, vol. 74 (1952), pp. 365-376.

[6] ———, *The Analytical Foundations of Celestial Mechanics*, 1941.

ON THE EXISTENCE OF RIEMANNIAN MANIFOLDS WHICH CANNOT CARRY NON-CONSTANT ANALYTIC OR HARMONIC FUNCTIONS IN THE SMALL.*

By PHILIP HARTMAN and AUREL WINTNER.

A function $f = f(x, y)$ on an open domain D of the real (x, y)-plane is said to be of class $C^n(\lambda)$, where $n \geqq 0$ and $0 \leqq \lambda \leqq 1$, if all partial derivatives $\partial^n f/\partial^m x \partial^{n-m} y$, where $0 \leqq m \leqq n$, exist and are continuous on D and satisfy a uniform Hölder condition of index λ on every compact subset of D. If the assumption of such an index λ is omitted, then $C^n(\lambda)$ reduces to the class $C^n = C^n(0)$. In particular, $C^0(1)$ is the class of functions satisfying a locally uniform Lipschitz condition, and $C^0 = C^0(0)$ is the class of all continuous functions, on D. In the questions to be dealt with below, there is no loss of generality in assuming that D is *schlicht*, simply connected and small, say

(1) $$D_a\colon\ x^2 + y^2 < a^2,$$

where $a > 0$ is arbitrarily fixed.

If $g_{11}, g_{12} = g_{21}, g_{22}$ are three real-valued functions of class $C^n(\lambda)$ on D and have the property that the matrix (g_{ik}) is positive definite at every point of D, then

(2) $$ds^2 = g_{11}dx^2 + 2g_{12}dxdy + g_{22}dy^2, \quad \text{where} \quad g_{ik} = g_{ik}(x, y),$$

will be called a $C^n(\lambda)$-metric (on D). It will be referred to as a C^n-metric if $\lambda = 0$, and as a continuous metric if it is a C^0-metric. If a in (1) is small enough and if two functions

(3) $$u = u(x, y), \qquad v = v(x, y)$$

are of class $C^{n+1}(\lambda)$, have a non-vanishing Jacobian on D_a and are normalized by

(4) $$(u(0, 0), v(0, 0)) = (0, 0),$$

then (3) has a unique inverse

(5) $$x = x(u, v), \qquad y = y(u, v),$$

* Received October 22, 1952.

where both functions (5) are of class $C^{n+1}(\lambda)$ on the neighborhood E_a of the point (4), if E_a denotes the image of D_a in the (u, v)-plane (and

$$(6) \qquad (x(0,0), y(0,0)) = (0,0)$$

is the image of the point (4) under the mapping (5)). Clearly, any such $C^{n+1}(\lambda)$-mapping transforms every $C^n(\lambda)$-metric (2) on D_a into a $C^n(\lambda)$-metric on E_a, say into

$$(7) \qquad ds^2 = h_{11}du^2 + 2h_{12}dudv + h_{22}dv^2, \quad \text{where} \quad h_{ik} = h_{ik}(u, v),$$

the coefficient matrix of (7) being given in terms of that of (2) by the requirement that

$$(8) \qquad (h_{ik}) = J(g_{ik})J' \text{ by virtue of (3) or (5)}$$

(in the sense of matrix multiplication), where J' denotes the transposed matrix of the Jacobian matrix J of (3) (in particular,

$$(9) \qquad \det h_{ik} = (\det J)^2 \det g_{ik},$$

where $\det g_{ik} > 0$ and $\det J \neq 0$). Conversely, if (2) and (7) are two $C^n(\lambda)$-metrics which become identical by virtue of a C^1-transformation (5) of non-vanishing Jacobian, then this $C^1(0)$-transformation must be of class $C^{n+1}(\lambda)$. This was proved in [5] for $\lambda = 0$ only, but the proof given there is valid for every λ.

Two continuous metrics, say (2) and (7), given on respective vicinities of points (6) and (4), are called isometric if these vicinities can be transformed into one another by C^1-transformations, (3) and (5), by virtue of which the two metrics become identical. Clearly, a continuous metric (7) of arbitrarily smooth coefficient functions (even the Euclidean metric $du^2 + dv^2$) is isometric with certain C^0-metrics (2) which are not C^1-metrics. It is therefore natural to consider the class, say C_0, of those continuous metrics, the "exactly continuous metrics," which are not isometric with any C^1-metric and, more generally, the class, say C_n, of those C^n-metrics which are not isometric with any C^{n+1}-metric.* The existence of an "exactly continuous metric" is

* There is a corresponding question for pairs of linear, instead of quadratic, differential forms, say

$$(2') \qquad a_1(x,y)\,dx + a_2(x,y)\,dx, \qquad\qquad (7') \qquad b_1(u,v)\,du + b_2(u,v)\,dv,$$

which are isometric in the sense that (2′) is identical with (7′) by virtue of some C^1-transformation (3) of non-vanishing Jacobian. Let a Pfaffian (2′) be called of class C^n if its coefficients $a_i(x,y)$ are of class C^n, and let (2′) be called of class C_n if it is of class C^n but is not isometric to any Pfaffian (7′) of class C^{n+1}. The existence of

not obvious at all. That such metrics exist will follow, as a corollary, from Theorem (i) below, if the latter is combined with the parenthetical assertion of Theorem (ii) below. Similarly, Theorem ($\mathrm{ii_n}$) below, when combined with ($\mathrm{ii_{n+1}}$), implies that the class C_n is not vacuous for any n.

If (2) is a continuous metric (on D_a) corresponding to which a transformation (3) (of D_a into an E_a) can be so chosen that h_{11} and h_{12} in (7) become identical and h_{12} becomes identically 0, then (2) is said to possess a conformal normal form (on D_a). The latter is characterized by the existence of a continuous function h satisfying

$$(10) \qquad ds^2 = h^2 \cdot (du^2 + dv^2),$$

where $h = \pm (\det g_{ik})^{\frac{1}{2}} \partial(u, v)/\partial(x, y)$ in view of (9), hence

$$(11) \qquad h = h(u, v) > 0$$

without loss of generality.

(i) *If* (2) *is a continuous metric on a circle* (1), *then, no matter how small the latter be chosen, there need not exist any transformation* (3), *of class C^1 and of non-vanishing Jacobian, which transforms* (2) *into a conformal normal form* (10).

The question as to the existence of such metrics (2) was raised in [14], p. 203, where it was shown (pp. 204-205) that the answer is in the negative if (2), instead of being definite as above ($\det g_{ik} > 0$), is indefinite ($\det g_{ik} < 0$) and, correspondingly, (10) is replaced by $ds^2 = h^2 \cdot (du^2 - dv^2)$. It should be noted that C^1-metrics cannot be admitted in (i); what is more, such metrics as comply with (i) cannot be $C^0(\lambda)$-metrics for any $\lambda > 0$. In fact, if $0 < \mu < \lambda \leqq 1$, then, according to Lichtenstein [10], there belongs to every $C^0(\lambda)$-metric (2) a transformation (3) which is of class $C^1(\mu)$ (hence of class C^1) along with its inverse (5), and transforms (2) into a conformal normal form (10). Thus the point of (i) is that Lichtenstein's arbitrarily small $\lambda > 0$ cannot be replaced by $\lambda = 0$.

Since every C^1-solution $u^* = u^*(u, v)$, $v^* = v^*(u, v)$ of the Cauchy-Riemann equations $u^*_u = v^*_v$, $u^*_v = -v^*_u$ is analytic, it is clear that if a continuous metric (2) has a conformal normal form in terms of the parameters u, v, then it will have a conformal normal form in some other parameters u^*, v^* if and only if $w^* = u^* + iv^*$ is an analytic function of

Pfaffians of class C_n is obvious only if $n > 0$. That $n = 0$ need not actually be excluded (i. e., that there exist "exactly continuous" Pfaffians), can readily be concluded from the result italicized in [14], Section 7, pp. 205-206.

$w = u + iv$ satisfying $dw^*/dw \neq 0$. But this universality (with regard to the choice of (2)) of the analytic functions of a complex variable depends on the assumption that (2) has *some* conformal normal form. And it turns out that there exist continuous metrics (2) for which this assumption is not satisfied in the neighborhood of any point; in other words, that *a positive definite, continuous, Riemannian metric* (2) *on the circle* $D_a\colon x^2 + y^2 < a^2$ *need not carry any non-constant analytic function* (on D_b, no matter how small $b < a$ be chosen).

The Cauchy-Riemann equations belonging to a continuous metric (2) are

$$(12) \qquad g v_x = g_{12} u_x - g_{11} u_y, \qquad g v_y = g_{22} u_x - g_{12} u_y$$

(Riemann, Beltrami; cf. [8], pp. 520-521, [1], pp. 126-127), where

$$(13) \qquad g = (\det g_{ik})^{\frac{1}{2}} > 0.$$

But (12) implies that the Jacobian $u_x v_y - u_y v_x$ cannot vanish identically unless

$$(14) \qquad u = \text{const.}, \qquad v = \text{Const.}$$

Hence it is clear that the last italicized statement is the substance of the following refinement of (i):

(i*) *There exist continuous metrics* (2) *for which the system* (12) *does not possess any solution distinct from* (14), provided that by "a solution (3)" is meant a pair of functions (3) for which the partial derivatives

$$(15) \qquad u_x, u_y; \qquad v_x, v_y$$

exist, satisfy (12) and are *continuous.*

Beltrami's first differential parameter belonging to (2) is $\nabla(u, u)$, where, if g is defined by (13),

$$(16) \qquad g^2 \nabla(u, z) = g_{22} u_x z_x - g_{12}(u_x z_y + u_y z_x) + g_{11} u_y z_y$$

(cf. [1], pp. 76-77). Hence Dirichlet's problem belonging to a continuous metric (2) is

$$(17) \qquad \min \iint_{[\Gamma]} g \nabla(u, u)\, dx dy \text{ when } u(x, y) = \phi \text{ on } \Gamma.$$

Here $[\Gamma]$ denotes the interior of a Jordan curve Γ contained in the domain (1) on which (2) is given as a continuous metric, and ϕ is a preassigned continuous function of position on Γ. What is sought for is a function $u(x, y)$ which is continuous on $[\Gamma] + \Gamma$, of class C^1 on $[\Gamma]$, equal to ϕ on Γ, and

such as to minimize the integral (17) with reference to all such functions $u(x, y)$. In view of Hilbert's method ([7], pp. 10-14 and pp. 15-37; cf. in particular the general comments referred to in the footnote on p. 11 and the italicized statement to which it belongs on p. 11), it seems to be of methodical interest that (i*) implies (and is substantially equivalent to) the following negative result:

(i bis) *There exist on* (1) *continuous metrics* (2) *corresponding to which the Dirichlet problem* (17) *has on* $[\Gamma]$ *no solution* $u(x, y)$ *of class* C^1 (*and continuous on* $[\Gamma] + \Gamma$) *with reference to any Jordan curve* Γ *contained in* (1) *and to any continuous non-constant boundary function* ϕ (if $\phi =$ const. on $[\Gamma]$, then $u(x, y) =$ const. is of course a solution of (17) on $[\Gamma] + \Gamma$).

It is essential that in this restatement (i bis) of (i*) no (x, y)-set of measure 0 is excluded from $[\Gamma]$; in this regard, cf. [12].

In order to prove (i bis), choose (2) as in (i*), suppose that the assertion of (i bis) is false for some Γ and some ϕ ($\neq$ const.) on Γ and denote by $u(x, y)$ the (or a) corresponding solution of (17). Then, if $z = z(x, y)$ is any function of class C^1 on $[\Gamma] + \Gamma$ satisfying $z \equiv 0$ on Γ, the value of the Dirichlet integral (17) is not less for $u + z$ than for u itself. Hence

$$\iint_{[\Gamma]} g\nabla(u, z)\, du dv = 0 \tag{18}$$

follows in the usual way (that is, from the bilinear character of the operator (16) and from the fact that the matrix of the bilinear form (16), being g times the matrix of (2), is positive definite). Insertion of (16) into (18) gives

$$\iint_{[\Gamma]} (az_x + bz_y)\, dx dy = 0 \tag{19}$$

if $a = a(x, y)$ and $b = b(x, y)$ are defined by

$$ga = g_{22}u_x - g_{12}u_y, \qquad gb = g_{11}u_y - g_{12}u_x. \tag{20}$$

Hence the functions a, b are continuous on the open set $[\Gamma]$ and, in view of the finiteness of the integral (17), Schwarz's inequality implies that a, b are of class L^2 (and therefore absolutely integrable) on $[\Gamma]$ or, since meas $\Gamma = 0$, on $[\Gamma] + \Gamma$. It follows therefore from (a trivial extension of) Haar's lemma [3], p. 2 (where a, b are supposed to be continuous on $[\Gamma] + \Gamma$), that the truth of (19) for each of the above-mentioned functions $z = z(x, y)$ implies the vanishing of

$$\int (ady - bdx), \tag{21}$$

where the integration path is any rectifiable Jordan curve contained in $[\Gamma]$. This means that the integral (21), when extended within $[\Gamma]$ from a fixed point (x_0, y_0) to a variable point (x, y), is a point function, say $v = v(x, y)$. But the function thus defined on $[\Gamma]$ has the partial derivatives $v_x = -b$, $v_y = a$, by (21). In view of (20) and (13), this means that $v(x, y)$ is of class C^1 on $[\Gamma]$ and satisfies (12). If this is compared with (i*), the assertion of (i bis) follows.

Theorem (i) will be paralleled by the following:

(ii) *If* (2) *is a C^1-metric on a circle* (1), *then there* (*exists a transformation* (3) *of class C^1 but*) *need not exist any transformation* (3) *of class C^2, of non-vanishing Jacobian, which transforms* (2) *into a conformal normal form* (10).

As mentioned after (i), the parenthetical (positive) assertion of (ii) is a corollary of Lichtenstein's theorem. In view of Riemann's mapping theorem and of the remarks made before (i), this assertion of (ii) is equivalent to the statement that all functions analytic on a circle of the ordinary complex $(u + iv)$-plane can be transferred to the circle (1) so as to become analytic functions with reference to the C^1-metric (2). In fact, the Cauchy-Riemann equations (12) are then satisfied. But the main (negative) assertion of (ii) is that the functions (3), which represent the real and imaginary parts of the analytic functions on (2), will become of class C^2 only in the trivial case (14), if the C^1-metric (2) is suitably chosen.

The analogue of the refinement (i*) of (i) is the following:

(ii*) *There are C^1-metrics on which there does not exist any non-constant harmonic function* (*although all analytic functions exist on every C^1-metric*), provided that by an harmonic function of a C^1-metric (2) is meant a function $u = u(x, y)$ for which the partial derivatives

$$u_x,\ u_y,\ u_{xx},\ u_{xy},\ u_{yx},\ u_{yy} \tag{22}$$

exist, are *continuous* and such as to satisfy the condition expressed by the identical vanishing of Beltrami's second differential operator, that is, by the partial differential equation

$$\{(g_{22}u_x - g_{12}y_y)/g\}_x + \{(g_{11}u_y - g_{12}u_x)/g\}_y = 0; \tag{23}$$

cf. [1], p. 127. Note that (23) is the (formal, whereas

(21 bis) $$\int_{\Gamma} (ady - bdy) = 0$$

with (20) is the unrestricted) integrability condition ($v_{xy} = v_{yx}$) of (12).

What belongs to (ii) in the same way as (i bis) belongs to (i) is the following circumstance:

(ii bis) *There exist on* (1) *metrics* (2) *of class* C^1 *corresponding to which the Dirichlet problem* (17) *belonging to a smooth Jordan curve* Γ *and a smooth boundary function* ϕ *can have a solution* $u(x, y)$ *of class* C^1, *although the Euler-Lagrange equation of* (17) *fails to have a non-constant solution of class* C^2.

In fact, the latter equation is (23), while the system of Haar ([3], pp. 16-17) belonging to (17) reduces to (12). Hence the last italicized statement follows from (ii*). A corresponding situation for the hyperbolic Euler-Lagrange equation $u_{xx} - u_{yy} = 0$ was pointed out by Hadamard [7], pp. 242-243, and for the elliptic (but inhomogeneous) Euler-Lagrange equation $u_{xx} + u_{yy} = f(x, y)$, where $f(x, y)$ is continuous, by Lichtenstein [9] (a corresponding example for the homogeneous equation $u_{xx} + u_{yy} + f(x, y)u = 0$ follows from [13], p. 733).

Both (i) and (ii) will be proved by choosing a suitable positive function $g = g(x, y)$ and placing $g_{11} = 1$, $g_{12} = 0$, $g_{22} = g^2$. Then (2) becomes

(24) $$ds^2 = dx^2 + g^2dy^2, \qquad (g > 0),$$

the g occurring in (24) is identical with the square root (13). The Cauchy-Riemann equations (12) simplify to

(25) $$v_x = -g^{-1}u_y, \qquad v_y = gu_x$$

if (24) is a C^0-metric (i. e., if g is continuous), and the Laplace equation (23) can be replaced by

(26) $$(gv_x)_x + (g^{-1}v_y)_y = 0$$

if (24) is a C^1-metric (i. e., if g is a function of class C^1). In fact, if the two linear equations (12) are solved with respect to u_x, u_y and, correspondingly, the integrability condition $v_{xy} = v_{yx}$ of (12), which is (23), is replaced by $u_{xy} = u_{yx}$, then what results in the case (24) of (2) is (26).

It will be clear from the proof of (ii) that (ii) can be generalized as follows:

(ii_n) *The assertions of* (ii) *remain true if the classes* C^1, C^2 *referred to in* (ii) *are replaced by* C^n, C^{n+1}, *respectively, where* n *is any positive integer.*

In view of (i), this holds for $n = 0$ also, except that what then corresponds to the parenthetical remark of (ii) must be omitted as meaningless (in fact, the functions (5) cannot be substituted into (2) if they are just of class C^0, i. e., continuous).

In the proof of (i), the following lemma (†) on inhomogeneous Cauchy-Riemann equations will be needed:

(†) *If* λ, μ *is a given pair of continuous functions on a circle*

$$D_a\colon\ x^2 + y^2 < a^2,$$

then the system

$$u_x - v_y = \lambda(x, y), \qquad u_y + v_x = \mu(x, y)$$

cannot possess any solution $u = u(x, y)$, $v = v(x, y)$ *of class* C^1 *on any circle* D_b *unless*

$$\lim_{\epsilon \to +0} \int_0^{2\pi} \left(\int_\epsilon^b r^{-1}(\lambda \cos 2\theta + \mu \sin 2\theta)\, dr \right) d\theta$$

exists as a finite limit (for every and/or some, sufficiently small, value $b > 0$ *of the upper limit of integration), where*

$$\lambda = \lambda(r \cos \theta, r \sin \theta), \qquad \mu = \mu(r \cos \theta, r \sin \theta).$$

It will be convenient to prove the above statements in the following order: (ii), (i), (†), (ii*), (i*).

Proof of (ii). Let $a < 1$ in (1) and define on (1) a continuous function h by placing

$$(27) \qquad h(x, y) = x^2/(r^2 \log r^2), \text{ if } 0 < r < a, \text{ and } h(0, 0) = 0,$$

where $r = (x^2 + y^2)^{\frac{1}{2}} \geqq 0$ (this is the function used by Petrini [11], p. 138, to show that Poisson's equation

$$(28) \qquad v_{xx} + v_{yy} = f(x, y)$$

need not have a solution $v(x, y)$ of class C^2 if $f(= h)$ is just continuous). It turns out (cf. [6], p. 136) that the function

$$(29) \qquad g(x, y) = 1 + \int_0^x h(t, y)\, dt$$

is of class C^1. Clearly, it is positive on (1) if a is sufficiently small. Thus (24) is a C^1-metric on the circle $D_a: x^2 + y^2 < a^2$ if a is small enough.

In order to prove (ii), it will first be shown that the case (29) of the Laplace equation (26) cannot have on D_a (for any sufficiently small a) any solution $v(x, y)$ of class C^2 satisfying

$$(30) \qquad v_x(0,0) = 1 \quad \text{and} \quad v_y(0,0) = 0.$$

This will be concluded by adapting the arguments used in [13], pp. 736-738 as follows:

Let $v = v(x, y)$ be any function of class C^2 satisfying (26). Then, for this v, (15) can be written in the form of a Poisson equation (28), in which $f = f(x, y)$ is

$$(31) \qquad f = f_1 + f_2 + f_3 + f_4,$$

if the four functions $f_j = f_j(x, y)$ are defined by

$$(32_1) \quad f_1 = (1 - g)v_{xx}; \qquad (32_2) \quad f_2 = (1 - g^{-1})v_{yy};$$

$$(32_3) \quad f_3 = g^{-2}g_y v_y; \qquad (32_4) \quad f_4 = -g_x v_x.$$

But if $f(x, y)$ is any function continuous (and, say, bounded) on a circle (1), then results of Zaremba [15] imply (cf. [13], p. 735) that the corresponding Poisson equation (28) has on (1) no continuous solution v (that is, there exists on (1) no continuous v possessing second derivatives ϕ_{xx}, ϕ_{yy} the sum of which is f) unless such a v is represented by the logarithmic potential, say $v^* = v^*(x, y)$, belonging to the density $(2\pi)^{-1}f(x, y)$; in which case every such v is of the form $v = v^* + w$, where w is a regular harmonic function on (1) (in the sense of the euclidean metric, i. e., $w_{xx} + w_{yy} = 0$). On the other hand, Petrini has shown ([11], pp. 131-134) that the partial derivative v^*_{xx} and/or v^*_{yy} of the logarithmic potential $v^*(x, y)$ will exist at $(x, y) = (0, 0)$ if and only if

$$(33) \qquad \lim_{\epsilon \to +0} \int_0^{2\pi} \left(\int_\epsilon^a r^{-1} f(r\cos\theta, r\sin\theta)\, dr \right) \cos 2\theta \, d\theta \ \text{exists}$$

(as a finite limit) for the function $f = f(x, y)$ defining v^* and occurring in (28). Hence, if (33) is satisfied by each of the three functions f_1, f_2, f_3 but is not satisfied by f_4, then the case (31) of (28) cannot have any solution v of class C^2 on (1), which means that the same is true of (26). Consequently, in order to prove that (26) has no solution v of class C^2 satisfying (30), it is

sufficient to show that the first three of the functions (32_1)-(32_4) which belong to such a v must, but the fourth of them cannot, satisfy condition (33).

First, since the function (29) is of class C^1 and becomes 1 at $(x, y) = (0, 0)$, it is clear that $1 - g = O(r)$, where $r = (x^2 + y^2)^{\frac{1}{2}} \to 0$. Hence (32_1), (32_2) and the continuity (in fact, just the boundedness) of v_{xx}, v_{yy} imply that $f_1 = O(r)$, $f_2 = O(r)$. Consequently, (33) is satisfied by $f = f_1$ and by $f = f_2$. On the other hand, since v is of class C^2, hence v_y of class C^1, the second of the assumptions (30) implies that $v_y = O(r)$. It follows therefore from (32_3), where $g \to 1$ and $g_y \to g_y(0, 0)$ as $r \to 0$, that $f_3 = O(r)$, and so (33) is satisfied by $f = f_3$ also. It remains to show that (33) is violated by $f = f_4$.

To this end, note that, since v_x is of class C^1, the first assumption in (30) and the definition (32_4) imply that $f_4 = (-1 + O(r))g_x$. It follows therefore from (29) that $f_4 = -h + O(r)$. Hence $f = f_4$ violates (33) if $f = h$ does. But the function $f = h$, defined by (27), is precisely Petrini's example of a continuous function violating (33); cf. [11], p. 138.

This proves that (26) cannot have a solution v of class C^2 satisfying (30). Hence, in order to prove (ii), it is sufficient to ascertain that the negation of the statement of (ii) implies the existence of such a v. Suppose therefore that (ii) is false with reference to the C^1-metric (24), defined by (29). Then, if a is small enough, there exists for the circle (1) a transformation (3) which is of class C^2 along with its inverse (5) and which transforms (24) into $du^2 + dv^2$ times a positive function of (u, v) (in view of (8), this positive function is of class C^1). Hence the C^2-mapping (3) transforms every function $t = t(u, v)$, which is a regular harmonic function (i. e., function of class C^2 satisfying the Laplace equation $t_{uu} + t_{vv} = 0$) in a vicinity of the point (4), into a function

$$\tau(x, y) \equiv t(u, v) \tag{34}$$

which, in a vicinity of the point (6), is of class C^2 and a solution $v = \tau$ of the Laplace equation (26) (the v in (26) will not be confused with the v in (3) or (5)). Since the regular harmonic function $t(u, v)$ is arbitrary in (34), and since the Jacobian of (3) does not vanish, it is clear that a $\tau(x, y)$ can be so chosen that the initial values $\tau_x(0, 0)$, $\tau_y(0, 0)$ of its first partial derivatives become preassigned numbers. Consequently, both (26) and (30) can be satisfied by a function $v = \tau(x, y)$ of class C^2. This contradiction completes the proof of (ii).

Proof of (i). Let $a < 1$ in (1) and define the continuous function h on D_a again by (27) (that is, by

$$h(x, y) = \tfrac{1}{2}(1 + \cos 2\theta)/\log r^2, \quad \text{where} \quad h(0, 0) = 0 \tag{35}$$

5

and $x = r\cos\theta$, $y = r\sin\theta$), but let (29) now be replaced by

$$(36) \qquad g(x,y) = 1 + h(x,y) \qquad (g(0,0) = 1).$$

Then, if a is small enough, $g(x,y)$ is positive, hence (24) is a continuous metric, on D_a. It will be shown that this metric has the property claimed by (i).

To this end, it will first be shown that the case (36) of the system (25) cannot possess on D_a any solution $u = u(x,y)$, $v = v(x,y)$ of class C^1 satisfying

$$(37) \qquad u_x(0,0) = 1, \qquad u_y(0,0) = 0.$$

In order to prove this, let u, v be any pair of functions which are of class C^1, and satisfy (25) identically in (x,y), on D_a. These two (x,y)-identities can be written in the form $u_x - v_y = \lambda$, $u_y + v_x = \mu$, where $\lambda = \lambda(x,y)$, $\mu = \mu(x,y)$ denote the functions defined by

$$(38) \qquad \lambda = (1-g)u_x, \qquad \mu = (1-g^{-1})u_y;$$

functions which are continuous on D_a, since u_x, u_y and $g\,(>0)$ are. It follows therefore from the statement of Lemma (†), italicized before (27), that the system (25) cannot have any solution of class C^1 satisfying (37) if the limit of the double integral, occurring in (†), fails to exist (as a finite limit) for the functions (38). Hence it is sufficient to show that the condition

$$\int_0^{2\pi} \Big(\int_\epsilon^b r^{-1}(\lambda\cos 2\theta + \mu\sin 2\theta)\,dr\Big)\,d\theta \to \infty, \text{ as } \epsilon \to +0,$$

is satisfied by the functions (38) and by some and/or all, sufficiently small, value of $b > 0$. In particular, it is sufficient to show that the integral is of the form

$$\int_\epsilon^b (r\log r^2)^{-1}(-\tfrac{1}{2}\pi + o(1))\,dr,$$

where the $o(1) = o(r)/r$ refers to $r \to 0$. Consequently, it is sufficient to ascertain that, if $x = r\cos\theta$ and $y = r\sin\theta$ in $\lambda = \lambda(x,y)$, $\mu = \mu(x,y)$, and if $r \to 0$, then, uniformly in θ,

$$\lambda = -\tfrac{1}{2}(1+\cos 2\theta)L + o(|L|) \text{ and } \mu = o(|L|), \text{ where } L = (\log r^2)^{-1}.$$

But the latter pair of relations is readily verified. In fact, (37) means that the (continuous) functions u_x, u_y are of the respective forms $1 + o(1)$, $o(1)$.

It follows therefore from (38) and (36) that λ is $-h + ho(1)$ and that μ is $1-(1+h)^{-1}$ times $o(1)$. Hence the pair of relations claimed by the last formula line follows from (35). This proves that (25) cannot have any solution of class C^1 satisfying (37).

In order to conclude from this that the continuous metric (24) defined by (36) has the property claimed in (i), suppose that it does not. The transformation supplied by this negation, a transformation of class C^1 and of non-vanishing Jacobian, can be used in the same way as, at the end of the proof of (ii), the corresponding transformation of class C^2 was used in conjunction with (34). In fact, it is sufficient to replace there the (real) harmonic functions t, occurring in (34), by the analytic functions $s + it$ which are regular at the origin of a complex plane. By virtue of the C^1-transformation of non-vanishing Jacobian, the real and imaginary parts of all these regular function elements become functions $u = u(x, y)$, $v = v(x, y)$ of class C^1 satisfying the Cauchy-Riemann equations (25) on a vicinity of $(x, y) = (0, 0)$. Since the class of these functions clearly contains pairs u, v for which u_x and u_y attain preassigned values at the origin, (37) is satisfied. But it was proved above that this is impossible.

This contradiction proves (i). But the proof depended on Lemma (†), which therefore remains to be proved.

Proof of (†). Let $[\Gamma]$ be the interior of a Jordan curve Γ which is sufficiently smooth (say, piecewise of class C^1), and let

$$u = u(x, y), \qquad v = v(x, y) \tag{39}$$

be a pair of real-valued functions which, on $[\Gamma]$, are of class C^1, uniformly continuous and such as to satisfy the pair of partial differential equations

$$u_x - v_y = \lambda(x, y), \qquad u_y + v_x = \mu(x, y), \tag{40}$$

where λ, μ are given functions which are uniformly continuous of $[\Gamma]$ (so that they, as well as the functions (39), possess continuous boundary values on Γ). Then the assertion of (†) is equivalent to the statement that, at every point (x, y) of $[\Gamma]$, the functions λ, μ must behave in such a way that, whenever $b > 0$ is small enough,

$$\lim_{\epsilon \to +0} \int_0^{2\pi} \left(\int_\epsilon^b \rho^{-1}(\lambda \cos 2\phi + \mu \sin 2\phi)\, d\rho d\phi \text{ exists} \right. \tag{41}$$

(as a finite limit), where the argument of both λ and μ is $(x + \rho \cos \phi, y + \rho \sin \phi)$. The proof (41) proceeds as follows:

If ∂ denotes the operator $(\ )_x + i(\ )_y$, then (40) can be written in the form $\partial w = \omega$, where $w = u + iv$ and $\omega = \lambda + i\mu$ (so that w is a complex-valued function of class C^1 of the real variables x, y on the domain $[\Gamma]$, on which both w and ω are uniformly continuous). Let $z = x + iy$ and $\zeta = \xi + i\eta$, and put

$$(42) \qquad 1/(\zeta - z) = G = G(x, y) = G(x, y; \xi, \eta)$$

if $z \neq \zeta$. Then $\partial G = 0$ is an identity in (x, y) for fixed $\zeta \neq z$. Hence the equation $\partial w = \omega$ can be written in the form $\partial(wG) = \omega G$. Consequently, an application of Green's theorem gives

$$\int_{\Gamma + C} wGdz = i \int\int_B \omega G dx dy,$$

where Γ is positively oriented, $C = C(\epsilon; \xi, \eta)$ denotes the negatively oriented circle of (small) radius ϵ about the point $\zeta = \xi + i\eta$ and $B = B(\epsilon; \xi, \eta)$ is the annular domain between Γ and C. But it is clear from (42) that

$$\int_C wGdz \rightarrow 2\pi i w(\xi, \eta) \text{ as } \epsilon \rightarrow 0.$$

Hence, by letting $\epsilon \rightarrow 0$ in the preceding relation also, it follows that

$$(43) \qquad w(x, y) + (2\pi i)^{-1} \int_\Gamma wGdz = (2\pi)^{-1} \int\int_{[\Gamma]} \omega G d\xi d\eta$$

holds for every point (x, y) of $[\Gamma]$. (Except for the notations, (43) is the same as formula (3) of Carleman [2], p. 473; cf. formulae (28) of Lichtenstein [10]).

It is clear from (42) that the function of (x, y) represented by the line integral on the left of (43) is regular analytic in $z = x + iy$ on $[\Gamma]$ (hence such as to annihilate the operator ∂). Since w is of class C^1 and satisfies $\partial w = \omega$ on $[\Gamma]$, it follows from (43) that w^* is of class C^1 (and satisfies $\partial w^* = \omega$) on $[\Gamma]$, if $w^* = w^*(x, y)$ denotes the expression on the right of (43).

With reference to a sufficiently small $b > 0$ (which can be kept fixed for every closed (x, y)-subset of $[\Gamma]$), let $D = D(x, y)$ denote the interior of the circle of radius b about (x, y). Then, since $w^*(x, y)$ is of class C^1, the contribution of D to the double integral $w^*(x, y)$ also is a function of class C^1. In particular, both partial derivatives f_x, f_y of the function

$$(44) \qquad f(x, y) = \int\int_D (\lambda + i\mu) G(\xi, \eta; x, y) d\xi d\eta,$$

where $D = D(x, y)$ and $\lambda + i\mu = \omega = \omega(\xi, \eta)$, must exist. Hence the same is true of the corresponding partial derivatives of $\Re f(x, y)$ and $\Im f(x, y)$. But if

$$\xi = x + \rho \cos \phi, \qquad \eta = y + \rho \sin \phi, \tag{45}$$

it is seen from (44) and (42) that the functions $\Re f(x, y)$, $-\Im f(x, y)$ are identical with

$$\int_0^{2\pi}\int_0^b \rho^{-1}(\lambda \cos \phi + \mu \sin \phi) d\rho d\phi, \qquad \int_0^{2\pi}\int_0^b \rho^{-1}(\lambda \sin \phi - \mu \cos \phi) d\rho d\phi \tag{46}$$

(where $\lambda = \lambda(\xi, \eta)$, $\mu = \mu(\xi, \eta)$, whilst ξ, η are given by (45)). Functions of the type

$$\int_0^{2\pi}\int_0^b e(\phi)\rho^{-1}\nu(x + \rho \cos \phi, y + \rho \sin \phi)\, d\rho d\phi,$$

where $e(\phi) = \cos \phi$ or $e(\phi) = \sin \phi$ and $\nu = \lambda$ or $\nu = \mu$, are precisely those treated by Petrini [11], pp. 128-130 (such functions are, in the main, partial derivatives of the first order of logarithmic potentials; cf. *ibid.*, p. 129). Petrini's proof ([11], pp. 131-132) of his theorem (pp. 132-133) implies that the partial derivative of $\Re f$ with respect to x (at the point (x, y)) will exist if and only if condition (41) is satisfied. This proves (†).

Proof of (ii*). In order to prove (ii*), it will be convenient to re-examine the proof of (ii). Consider the case (24) of (2), where g is the function of class C^1 given by (29). Suppose, if possible, that the corresponding Laplace-Beltrami equation (26) has, on a vicinity of $(x, y) = (0, 0)$, a solution $v = v(x, y)$ of class C^2 satisfying

$$v_x^2(0, 0) + v_y^2(0, 0) \neq 0. \tag{47}$$

Since (26) is the integrability condition of (25), it follows that (25) defines a function $u = u(x, y)$, unique up to an additive constant. It is clear that u is of class C^2, since v is of class C^2 and g is of class C^1. It also follows from (25) that $u_x v_y - u_y v_x = g^{-1} v_x^2 + g v_y^2$, which, in view of (47) and $g(0, 0) = 1 \neq 0$, implies that $\partial(u, v)/\partial(x, y)$ does not vanish at, hence near, $(x, y) = (0, 0)$. This contradicts the proof of (ii). Consequently, any C^2-solution of (26) on a vicinity of $(x, y) = (0, 0)$ fails to satisfy (47).

The proof of (ii), and hence the last conclusion, is based essentially on two properties of g, namely, that

$$g(0, 0) = g^{-1}(0, 0) = 1 \tag{48}$$

and that, for some $b > 0$,

$$(49) \qquad \lim_{\epsilon \to 0} \int_0^{2\pi} \int_\epsilon^b r^{-1} \cos 2\theta \; h(r \cos \theta, r \sin \theta) \, dr d\theta \text{ does not exist}$$

as a finite limit, where $h = g$. If the 1 in (29) is replaced by a $c(> 0)$, so that (48) does not hold, then, in the proof of (ii), the Poisson equation (28) becomes replaced by $c^2 v_{xx} + v_{yy} = f$. This can be reduced to a Poisson equation by a change of the independent variables $(cx, y) \to (x, y)$. The condition corresponding to (49) holds if, for example,

$$(50) \qquad \lim_{\epsilon \to 0} \int_0^{2\pi} \int_\epsilon^b r^{-1} \cos 2\theta \; h(cr \cos \theta, r \sin \theta) \, dr d\theta = -\infty.$$

In the proof of (ii*), it is sufficient to observe that (50) holds for all c near 1. In order to see this, note that the integral occurring in (50) is an absolutely convergent double integral over a region bounded by the circles $r = \epsilon$ and $r = b$, where $r^2 = x^2 + y^2$. If the inner boundary is replaced by the ellipse $c^2 x + y^2 = \epsilon^2$, the domain of integration is changed so that the area of the region added or subtracted is $| 1 - c^{-1} | \pi \epsilon^2$, while the integrand in this region is $O(\epsilon^{-1})/\log \epsilon$ as $\epsilon \to 0$; cf. (27). Hence the difference between the two integrals is $O(\epsilon)/\log \epsilon = o(1)$, as $\epsilon \to 0$. Thus, in proving (50), the inner and, of course, the outer boundary can be supposed to be changed to the ellipses $c^2 x^2 + y^2 = \epsilon^2$, $c^2 x^2 + y^2 = b^2$, respectively. The integral to be considered is then

$$c \int_0^{2\pi} \int_\epsilon^b r^{-1} (\cos^2 \theta - c^2 \sin^2 \theta)(\cos^2 \theta + c^2 \sin^2 \theta)^{-2} h(r \cos \theta, r \sin \theta) dr d\theta,$$

or, in view of (27),

$$c \int_0^{2\pi} \left(\int_\epsilon^b r^{-1} \log r^2 \, dr \right) \cos^2 \theta (\cos^2 \theta - c^2 \sin^2 \theta)(\cos^2 \theta + c^2 \sin^2 \theta)^{-2} d\theta.$$

Hence, (50) holds whenever

$$\int_0^{2\pi} \cos^2 \theta (\cos^2 \theta - c^2 \sin^2 \theta)(\cos^2 \theta + c^2 \sin^2 \theta)^{-2} d\theta > 0.$$

This is the case if $c = 1$ and hence if c is sufficiently near 1.

The proof of (ii*) can now be completed along the lines of [13], pp. 736-737, by the usual type of "Lebesgue construction," as follows:

Let $a > 0$ be chosen in (1) so small that the function g defined by (29)

is positive (and of class C^1) on (1). Let $(x_1, y_1), (x_2, y_2), \cdots$ be a sequence of distinct points of $D_{\frac{1}{2}a}$ dense on $D_{\frac{1}{2}a}$ and let δ, η be fixed positive numbers. For $n = 1$ and $n > 1$, respectively, put $A_n = 1$ and

$$(51) \qquad A_n^{-1} = 2^n \max_{1 \leqq k < n,\, |1-c| \leqq \eta} \int\int_E r^{-1} \,|\, h(x_k - x_n + cr\cos\theta, y_k - y_n + r\sin\theta) - h(x_k - x_n, y_k - y_n) \,|\, dr d\theta,$$

where $E = E(k, n, c)$ indicates the (r, θ)-set for which the argument of h is a point of D_a. Clearly, $0 < A_n <$ const. 2^{-n} and the series

$$(52) \qquad g^*(x, y) = \delta \sum_{n=1}^{\infty} A_n g(x - x_n, y - y_n)$$

defines a function which is positive and of class C^1 on $D_{\frac{1}{2}a}$. In addition, (52) can be differentiated term-by-term. Let $\eta > 0$ be chosen so that (50) holds for every c on the range $|\, c - 1 \,| \leqq \eta$ and let $\delta > 0$ be chosen so that $g^*(0, 0) = 1$. Then $|\, g^*(x, y) - 1 \,| \leqq \eta$ if (x, y) is on D_a for some sufficiently small $a (< \frac{1}{2}a)$.

Let $c = c_k = g^*(x_k, y_k)$ if (x_k, y_k) is on D_a; so that $|\, c - 1 \,| \leqq \eta$. If g is replaced by g^* in (26), the resulting equation cannot have, in a vicinity of $(x, y) = (x_k, y_k)$, a solution $v = v(x, y)$ of class C^2 satisfying

$$v_x^2(x_k, y_k) + v_y^2(x_k, y_k) \neq 0$$

if, for some $b > 0$,

$$(53) \qquad \lim_{\epsilon \to 0} \int_0^{2\pi} \int_\epsilon^b r^{-1} \cos 2\theta \;\; g^*_x(x_k + cr\cos\theta, y_k + r\sin\theta)\, dr d\theta = -\infty.$$

In order to verify (53), note that if the series (52) is differentiated term-by-term with respect to x and if the result is substituted into (53), then the integral occurring in (53) can be written as the sum of k integrals and a remainder term. The latter is majorized, uniformly in ϵ, by

$$\sum_{n=k+1}^{\infty} A_n I_n \leqq \text{const.} \sum_{n=k+1}^{\infty} 2^{-n},$$

where $I_n = I_n(k)$ is the integral occurring in (51). Of the first k integrals, the first $k - 1$ tend, as $\epsilon \to 0$, to finite limits, since $(x_k, y_k) \neq (x_n, y_n)$ for $k < n$, while the k-th integral is, up to a constant positive factor, the integral occurring in (50) and tends therefore to $-\infty$, as $\epsilon \to 0$.

Hence, if (26), where g is replaced by g^*, has a solution $v = v(x, y)$ of class C^2 on a subdomain of D_a, then $v_x(x_k, y_k) = v_y(x_k, y_k) = 0$ for every (x_k, y_k) on the domain of existence of v. Since $(x_1, y_1), (x_2, y_2), \cdots$ contains a subsequence dense on D_a, it follows that $v_x \equiv v_y \equiv 0$. Thus (ii*) is proved.

Proof of (i*). The proof of (i) implies that if g in the Cauchy-Riemann equations (25) is defined by (36) and if $u = u(x, y)$, $v = v(x, y)$ is a C^1-solution of (25) in a vicinity of $(0, 0)$, then $u_x = u_y = v_x = v_y = 0$ at $(x, y) = (0, 0)$. This conclusion depended on (48) and (49), where $g = 1 + h$. If the 1 in the last relation is replaced by a $c (> 0)$, then the corresponding equations (25) can be written as a system occurring in (†) after the change of independent variables $(cx, y) \rightarrow (x, y)$. The condition corresponding to (49) is implied by (50). Thus, it is clear that the proof of (i*) can be completed by arguments analogous to those used in the proof of (ii*).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] L. Bianchi, *Lezioni di geometria differenziale*, 3rd ed. (1927), vol. 1, Part I.

[2] T. Carleman, "Sur les systèmes linéaires aux dérivées partielles du premier ordre à deux variables," *Comptes Rendus*, vol. 197 (1933), pp. 471-474.

[3] A. Haar, "Über die Variation der Doppelintegrale," *Journal für die reine und angewandte Mathematik*, vol. 149 (1919), pp. 1-18.

[4] J. Hadamard, *Selecta* (1935).

[5] P. Hartman, "On unsmooth two-dimensional Riemannian metrics," *American Journal of Mathematics*, vol. 74 (1952), pp. 215-226.

[6] ——— and A. Wintner, "On the curvatures of a surface," *ibid.*, vol. 75 (1953), pp. 127-141.

[7] D. Hilbert, *Gesammelte Abhandlungen*, vol. 3 (1935).

[8] F. Klein, *Gesammelte mathematische Abhandlungen*, vol. 3 (1923).

[9] L. Lichtenstein, "Über das Verschwinden der ersten Variation bei zweidimensionalen Variationsproblemen," *Mathematische Annalen*, vol. 69 (1910), pp. 514-516.

[10] ———, "Zur Theorie der konformen Abbildung. Konforme Abbildung nichtanalytischer, singularitätenfreier Flächenstücke auf-ebene Gebiete," *Bulletin International de l'Académie des Sciences de Cracovie*, ser. A, 1916, pp. 192-217.

[11] H. Petrini, "Les dérivées premières et secondes du potentiel logarithmique," *Journal de Mathématiques*, ser. 6, vol. 5 (1909), pp. 127-223.

[12] J. Schauder, "Sur les équations linéaires du type elliptique à coefficients continus," *Comptes Rendus*, vol. 199 (1934), pp. 1366-1368; "Sur les équations quasi linéaires du type elliptique à coefficients continus," *ibid.*, pp. 1566-1568.

[13] A. Wintner, "On the Hölder restrictions in the theory of partial differential equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 731-738.

[14] ———, "On isometric surfaces," *ibid.*, vol. 74 (1952), pp. 198-214.

[15] S. Zaremba, "Contribution à la théorie d'une équation fonctionnelle de la physique" *Rendiconti del Circolo Matematico di Palermo*, vol. 19 (1905), pp. 140-150.

ON THE SINGULARITIES IN NETS OF CURVES DEFINED BY DIFFERENTIAL EQUATIONS.*

By Philip Hartman and Aurel Wintner.

1. Various questions in the differential geometry of surfaces are known to depend on a differential equation of the form

$$adx^2 + 2bdxdy + cdy^2 = 0, \tag{1}$$

where a, b, c are continuous functions of (x, y). If the discriminant $ac - b^2$ is negative at a point (x_0, y_0), then (1) is equivalent (after a rotation of the (x, y)-plane) to two non-singular differential equations of the form $dy/dx = f(x, y)$, where (x, y) is confined to a sufficiently small vicinity of (x_0, y_0). If $ac - b^2$ is negative near (x_0, y_0) and vanishes at (x_0, y_0) but a, b, c do not vanish simultaneously, then (1) is still "equivalent" to two non-singular differential equations. In the latter case, the solution curves of (1) passing through (x_0, y_0) have the same tangent, whereas there are two distinct such directions when $ac - b^2$ is negative at (x_0, y_0) also.

The first part of the present paper will be concerned with (1) in the case of an isolated singular point (x_0, y_0). By this is meant that $ac - b^2$ is negative near (x_0, y_0) but vanishes at (x_0, y_0) in such a way that both a and c, and therefore all three coefficients of (1), vanish at (x_0, y_0). The results will then be applied to the lines of curvature at an isolated umbilical point (which can be either a spherical or a flat point) of a surface (Section **11**), and also to the asymptotic lines at an isolated umbilical point (which must be a flat point) of a surface of non-positive curvature (Section **12**).

2. On a vicinity of $(x, y) = (0, 0)$, let the coefficient functions of (1) be continuous functions of the form

$$a = \alpha_1 x + \beta_1 y + f_1, \quad b = \alpha_2 x + \beta_2 y = f_2, \quad c = \alpha_3 x + \beta_3 y + f_3, \tag{2}$$

where α_k, β_k are six constants and each of the three functions $f_k = f_k(x, y)$ satisfies

$$f_k(x, y) = o(r), \qquad (k = 1, 2, 3), \tag{3}$$

* Received May 16, 1952.

as $r \to 0$, where $r = (x^2 + y^2)^{\frac{1}{2}}$. Supose further that

$$(4) \qquad ac - b^2 \leqq 0 \text{ according as } x^2 + y^2 \geqq 0.$$

By a solution path of (1) will be meant a set of points (x, y) which can be represented as a locus $x = x(\tau)$, $y = y(\tau)$, where $x(\tau)$, $y(\tau)$ are continuously differentiable functions satisfying (1) on a τ-interval and neither $(x(\tau), y(\tau))$ nor $(dx/d\tau, dy/d\tau)$ becomes the vector $(0, 0)$ at any point of that interval. The set of the solution paths of (1) contained in a vicinity of a point $(x^0, y^0) \neq (0, 0)$ can be divided into two well-distinguishable systems, since, in a vicinity of $(x^0, y^0) \neq (0, 0)$, the differential equation (1) splits into two non-singular differential equations the solution paths of which, in view of (4), have two distinct tangents at (x^0, y^0). It follows that, despite the singularity of (1) at $(0, 0)$, *all* solution paths of (1) can be divided into two distinct systems, say S_1 and S_2.

By a solution path of (1) reaching to the origin will be meant a solution path $x = x(\tau)$, $y = y(\tau)$ defined on an interval of the form $0 \leqq \tau < \tau_0$ $(\leqq \infty)$ in such a way that

$$(5) \qquad (x(\tau), y(\tau)) \to (0, 0) \text{ as } \tau \to \tau_0.$$

If r and θ denote polar coordinates,

$$(6) \qquad r = (x^2 + y^2)^{\frac{1}{2}}, \qquad \theta = \arctan y/x,$$

then, since $(x, y) \neq (0, 0)$ on a solution path, (6) defines continuously differentiable functions $r = r(\tau) > 0$, $\theta = \theta(\tau)$ when $x = x(\tau)$, $y = y(\tau)$ is a solution path.

In terms of the six constants occurring in (2), define three linear trigonometric forms by

$$(7) \qquad L_k(\theta) = \alpha_k \cos\theta + \beta_k \sin\theta, \qquad (k = 1, 2, 3),$$

and then a quadratic trigonometric polynomial by

$$(8) \qquad Q(\theta) = L_2^2(\theta) - L_1(\theta) L_3(\theta)$$

and a cubic trigonometric polynomial by

$$(9) \qquad M(\theta) = L_1(\theta)\cos^2\theta + 2L_2(\theta)\sin\theta\cos\theta + L_3(\theta)\sin^2\theta.$$

3. The following theorem will first be proved:

(*) *On a vicinity of* (x, y), *let* $a(x, y)$, $b(x, y)$, $c(x, y)$ *be continuous functions satisfying* (2), (3), (4) *and the following three conditions*:

(7 bis) $L_1(\theta), L_2(\theta), L_3(\theta)$ have no common zero;

(8 bis) $Q(\theta) \not\equiv 0$; (9 bis) $M(\theta) \not\equiv 0$.

Then, if S_1 and S_2 are defined as in Section 2,

(I) *each of the systems S_1, S_2 of solution paths of* (1) *contains at least one solution path satisfying* (5);

(II) *there belongs to every solution path of* (1) *satisfying* (5) *an angle θ_0 such that both*

(10) $$\theta \to \theta_0, \quad \text{where} \quad \theta = \arc\tan y/x,$$

and

(11) $$\phi \to \theta_0 \pmod{\pi}, \quad \text{where} \quad \phi = \arc\tan dy/dx,$$

hold, as $(x, y) \to (0, 0)$.

The proof of this theorem (*) will be based on the results of [3].

Remark 1. In view of (3), it is clear from (4) that

(12) $$Q(\theta) \geqq 0.$$

But if (12) is strengthened to

(13) $$Q(\theta) > 0$$

(for all θ) or, what is the same thing, if (4) is strengthened to

(14) $$ac - b^2 < -\text{const.}\ r^2 < 0, \quad \text{where} \quad x^2 + y^2 = r^2 \neq 0,$$

then (7 bis) and (9 bis) are automatically satisfied; cf. (7), (8), (9). Accordingly,

(14*) (7 bis), (8 bis), (9 bis) are implied by (14).

Remark 2. If the $o(r)$-terms (3) are neglected in (2), then the coefficients of (1) become the linear forms $\alpha_k x + \beta_k y$. Hence it is seen from (7) and (9) that assumption (9 bis) of the italicized Theorem (*) can be formulated as follows: The "linear" approximation to (1) is *not* of the form

$$\beta_1 y dx^2 - (\beta_1 x + \alpha_3 y) dx dy + \alpha_3 x dy^2 = 0.$$

On the other hand, every half-line issuing from the origin of the (x, y)-plane is readily seen to be a solution path of every differential equation of the latter form. If the analogy of the corresponding differential equations of first order

$$(a_{11}x + a_{12}y)\,dy - (a_{21}x + a_{22}y)\,dx = 0$$

is considered, every half-line issuing from the point $(0,0)$ will be a solution path if and only if the binary matrix (a_{ik}) (which should not be the zero matrix) has a multiple characteristic number but no multiple elementary divisor. Hence, in view of a counter-example which is known in this case for the (non-linear) first degree analogue of (non-linear) differential equation (1) of second degree (cf. [3], p. 123), assumption (9 bis) of (*) seems to be indispensable for the truth of (*).

In this connection, it is worth emphasizing that the assumptions of (*) do not preclude the case in which, when the terms (3) of (2) are omitted, (1) factors into

$$dx \cdot \{(\beta x + ay)\,dx - (ax - \beta x)\,dy\} = 0,$$

where both constants a, β can be distinct from 0. But then the vanishing of the second factor is a linear differential equation $\{\quad\} = 0$ for which the point $(0,0)$ is a vortex.

4. If $\theta = \theta_i$ and $\theta = \theta_i + \pi$, where $1 \leqq i \leqq h$, denote the $2h$ distinct $(\operatorname{mod} 2\pi)$, real roots of $M(\theta) = 0$ (so that $h = 1, 2$ or 3), then it will be clear from the proof below that the assertions of (*) above can be amplified as follows: The numeration of these roots can be chosen in such a way that, if S_1 and S_2 denote the systems defined above (before formula (5), in Section 2) and if the limit θ_0 occurring in (10)-(11) is not a zero of $Q(\theta)$, then θ_0 will be $(\operatorname{mod} 2\pi)$ a root θ_i or a root $\theta_i + \pi$ according as the solution path considered in (10)-(11) belongs to S_1 or to S_2. In addition, if $\theta = \theta_0$ is a root of odd order of $M(\theta) = 0$, then one of the two systems S_1, S_2 must contain a solution path satisfying (5) and (10), while the other system must contain a solution path satisfying (5) and what results if θ_0 is replaced by $\theta_0 + \pi$ in (10). Finally, if two solution paths reaching to the origin belong to one and the same θ_0 $(\operatorname{mod} 2\pi)$ in (10), and if $Q(\theta_0) \neq 0$, then both paths are contained in one and the same S_j $(j = 1, 2)$.

Since (*) requires only the continuity of the functions (3), there is assumed no local uniqueness (say a Lipschitz condition) for the solution paths of either system S_j at points (x, y) distinct from $(0, 0)$. But this generality will not be the only new aspect in the proof below, since the literature consulted fails to contain a proof of (*) even for the case in which the functions (2) are analytic and (14) holds. In fact, the situation is as follows:

For "generic" values of the six constants in (7), the assertions of (*)

were obtained by Picard [5], p. 224 (cf. Liebmann's report [4]) under the assumption that the functions (3) are regular power series about the origin (actually, somewhat less is used *loc. cit.*). A careful perusal of Picard's considerations shows however that this proof is wrong (even if the functions (3) are polynomials). For, in order to show that in the case of (14) no solution path reaching to the origin is a spiral (i. e., that $|\theta(\tau)| \rightarrow \infty$ cannot take place as $\tau \rightarrow \tau_0$), Picard assumes (*loc. cit.*, p. 221) that one of the two systems S_j, say S_1, contains two solutions reaching to the origin in such a fashion that one of the solutions satisfies (10) for some θ_0 and the other solution satisfies the relation which results from (10) if that θ_0 is increased by π. Actually, such an assumption cannot be made; in fact, as observed above, when $Q(\theta_0) \neq 0$, it is impossible that the latter of the solution paths be in S_1, since the former is assumed to be in S_1. Incidentally, if this result is granted, the error could be eliminated by using (*via* analyticity or less) the local uniqueness of solution paths passing through any point distinct from $(0, 0)$, since then it is easy to see that no solution path reaching to the origin can be a spiral.

An objection can also be made to the passage in which Picard assumes (*loc. cit.*, p. 219) that $\alpha_1\beta_3 - \alpha_3\beta_1 \neq 0$ holds for the constants occurring in the above relations (2). This assumption is *always* violated in the principal application made by Picard (*loc. cit.*, p. 225) of his result, namely, in the case in which (1) is the equation of the lines of curvature on a surface on which $(x, y) = (0, 0)$ is an isolated umbilical point. In fact, $\alpha_1 = -\alpha_3$ and $\beta_1 = -\beta_3$ always hold in this geometrical problem; so that $\alpha_1\beta_3 - \alpha_3\beta_1 = 0$ for any choice of the coordinate axes in the (x, y)-plane.

5. Assumption (7 bis) of (*) is violated if and only if there exist a homogeneous linear trigonometric form

$$(15) \qquad L(\theta) = \alpha \cos\theta + \beta \sin\theta \not\equiv 0 \qquad (\text{i. e., } \alpha^2 + \beta^2 \neq 0)$$

and three constants c_k satisfying

$$(16) \qquad L_k(\theta) = c_k L(\theta), \qquad (k = 1, 2, 3),$$

which implies, by (8) and (9), that

$$(17) \qquad Q(\theta) = (c_2{}^2 - c_1c_3)L^2(\theta)$$

and

$$(18) \qquad M(\theta) = (c_1 \cos^2\theta + 2c_2 \cos\theta \sin\theta + c_3 \sin^2\theta)L(\theta),$$

where, in view of (12) and (8 bis),

$$(18\text{ bis}) \qquad c_2^2 - c_1 c_3 > 0.$$

We were unable to decide whether or not (*) remains true in this case, that is, when assumption (7 bis) is omitted. We shall prove however, by a method which combines that proving (*) with that of "the curves of zero slope" (cf. [8]), that *if the coefficient functions of* (1) *are of class* C^1 (*instead of being, as in* (*), *just continuous*), then, if condition (7 bis) of (*) fails to hold, the assertions of (*) remain true at least in the following sense:

(§) *Assertion* (I) *of* (*) *remains true, as does that part of assertion* (II) *which concerns* (10); *the remaining part of* (II), *that which concerns* (11), *is true at least if the limit* θ_0 *occurring in* (10) *is not a zero of* (15).

The proof of this variant of (*), being made quite elaborate by the necessity of involving the "curves of zero slope," will be deferred to Section **15**. On the other hand, it will be shown in Section **14** that, even when the C^1-assumption of the variant is omitted, statement (§), the last italicized statement, holds if (7 bis) is replaced by the assumption that all (real) zeros of (9) are simple (i. e., if $M(\theta)$ and $dM(\theta)/d\theta$ do not vanish simultaneously). In contrast to the former variant of (*), the latter variant of (*) can be proved with not more effort than (*) itself.

6. If the (x, y)-plane is rotated about $(0, 0)$ by any fixed angle, say by θ^*, the form of the equations (1)-(3) remains unchanged. In fact, if $L_k^*(\theta)$, $Q^*(\theta)$, $M^*(\theta)$ denote the trigonometric polynomials by which (7), (8), (9) become replaced after the rotation

$$(19) \qquad (x, y) \rightarrow (x \cos\theta^* - y \sin\theta^*, \qquad x \sin\theta^* + y \cos\theta^*),$$

then it is readily verified from (1)-(3) and (6)-(7) that, in vector and matrix notations,

$$(20_1) \qquad \begin{pmatrix} L_1^*(\theta) \\ L_2^*(\theta) \\ L_3^*(\theta) \end{pmatrix} = \begin{pmatrix} \cos^2\theta^* & \sin 2\theta^* & \sin^2\theta^* \\ -\tfrac{1}{2}\sin 2\theta^* & \cos 2\theta^* & -\tfrac{1}{2}\sin 2\theta^* \\ \sin^2\theta^* & -\sin 2\theta^* & \cos^2\theta^* \end{pmatrix} \begin{pmatrix} L_1(\theta+\theta^*) \\ L_2(\theta+\theta^*) \\ L_3(\theta+\theta^*) \end{pmatrix}$$

and therefore, from (8) and (9),

$$(20_2) \qquad Q^*(\theta) = Q(\theta+\theta^*); \qquad\qquad (20_3) \qquad M^*(\theta) = M(\theta+\theta^*).$$

While a rotation (19) leaves the zeros of $M(\theta)$ invariant in the sense of (20_3), the positions of the zeros of the functions $L_k(\theta)$ do not remain invariant relative to the position of the zeros of $M(\theta)$; cf. (20_1). This is the reason for the possibility of the following consideration.

It is clear from (7) that, when θ is fixed, $L_k^*(\theta-\theta^*)$ is a quadratic form in $(\cos\theta^*, \sin\theta^*)$. If $k=3$, the coefficients of this quadratic form are seen to be $L_3(\theta)$, $-L_2(\theta)$, $L_1(\theta)$. Hence the assumption (7 bis) means that for no fixed θ will $L_3(\theta-\theta^*)$ vanish identically in θ^*. In view of (8), (8 bis) and (20_1), the function $L_3^*(\theta-\theta^*)$ of θ^* has, for a fixed θ, exactly one zero or exactly two zeros (mod π) according as θ is or is not a zero of the quadratic form $Q(\theta)$ in $(\cos\theta, \sin\theta)$.

Suppose that θ_0 is a zero of $M(\theta)$. Thus, (20_3) shows that $\theta=\theta_0-\theta^*$ is a zero of $M^*(\theta)$. Hence, in view of the preceding remarks on $L_3^*(\theta-\theta^*)$, it is seen from (9 bis) that, when θ^* is suitably chosen, $L_3^*(\theta)$ and $M^*(\theta)$ will not have a common zero, which means that $L_3^*(\theta)\neq 0$ holds at $\theta=\theta_0-\theta^*$ whenever $M(\theta_0)=0$. It also follows that if ϑ is any preassigned angle, then θ^* in (19) can be chosen in such a way that $L_3^*(\theta)\neq 0$ whenever $M^*(\theta+\vartheta)=0$, that is, whenever $\theta=\theta_0-\theta^*-\vartheta$ and $M(\theta_0)=0$.

Consequently, if the rotation (19) is suitably chosen and then the asterisks are omitted, it follows that there is no loss of generality in assuming that

$$L_3(\theta_0)\neq 0 \text{ if } M(\theta_0)=0 \tag{21}$$

and that, with reference to a preassigned ϑ,

$$L_3(\theta_0+\vartheta)\neq 0 \text{ if } M(\theta_0)=0. \tag{22}$$

The hypothesis of (21)-(22) is always satisfied by some θ_0. In fact, since (9) is a cubic form in $(\cos\theta, \sin\theta)$, it must have a (real) zero, say θ_0. In adidtion, (9 bis) shows that $M(\theta)$ will change sign at θ_0 if θ_0 is suitably chosen.

7. Starting with the coefficients of (1), which (for small x^2+y^2) are continuous functions satisfying (4), consider either of the binary differential systems

$$x'=c, \qquad y'=-b+(-1)^j(b^2-ac)^{\frac{1}{2}}, \tag{23_j}$$

where $j=1,2$, the prime denotes differentiation with respect to a variable t which does not occur explicitly in (23_j), and the exponent $\frac{1}{2}$ refers to the non-negative determination of the square root.

In contrast to the definition of a solution path of (1), given in Section **2**, where neither $(x(\tau), y(\tau))=(0,0)$ nor $(dx(\tau)/d\tau, dy(\tau)/d\tau)=(0,0)$ has been allowed, let a solution path $(x(t), y(t))$ of either system (23_j) be defined so as to exclude $(x(t), y(t))=(0,0)$ for every t, without excluding $(x'(t), y'(t))=(0,0)$. Clearly, (6) and every solution path of (23_j) deter-

mine two continuously differentiable functions $r = r(t) > 0$, $\theta = \theta(t)$ (with $0 \leqq \theta(t^0) < 2\pi$ at a given t^0). If a solution path of (23_j) is given for $0 \leqq t < t_0 (\leqq \infty)$ or $0 \geqq t > t_0 (\geqq -\infty)$ and if it satisfies

$$(24) \qquad (x(t), y(t)) \to (0, 0) \text{ as } t \to t_0,$$

then it will be called a solution path of (23_j) reaching to the origin.

The pair of alternative systems (23_1)-(23_2) of first order is *formally* equivalent to the single equation (1) of second order. But from the point of view of *solution paths,* the equivalence is not evident at all. In fact, a solution path of (1), as defined in Section **2**, depends on the idea of a (locally) Jordan arc, of class C^1, on which the parameter τ is in the main the arc length, whereas the t in (23_j), where $' = d/dt$, is committed by the assignment of the slope functions of both $z = x(t)$ and $z = y(t)$ in a (t, z)-plane. Correspondingly, when proving that every solution path of (1) satisfying (5) can be thought of as a solution path of either (23_1) or (23_2) and vice versa, one meets the actual difficulty at the points (x, y) at which $(x'(t), y'(t)) = (0, 0)$ by virtue of (23_j).

8. Let the functions F and G_1, G_2 be defined by

$$(25) \qquad F(\theta) = L_3(\theta)$$

and

$$(26_j) \qquad G_j(\theta) = -L_2(\theta) + (-1)^j Q^{\frac{1}{2}}(\theta),$$

where $Q^{\frac{1}{2}}$ denotes the non-negative square root of (8). Then it is clear from (2)-(3) and (7)-(8) that $1/r$ times the functions on the right of the equations (23_j) tend, uniformly in θ, to the limits $F(\theta)$ and $G_j(\theta)$, as $r \to 0$. The set of θ-values on which $F^2 + G_j^2$ vanishes is contained in the set of θ-values on which F vanishes, and the latter set is a sequence of the form $\theta = \theta^0 + n\pi$ $(n = 0, \pm 1, \pm, \cdots)$, by (25) (and since L_3 is a homogeneous linear trigonometric polynomial which, in view of (21), does not vanish identically).

If

$$(27_j) \qquad J_j(\theta) = G_j(\theta) \cos\theta - L_3(\theta) \sin\theta, \qquad (L_3 = F),$$

then it is seen from (9) that

$$(28) \qquad M(\theta) L_3(\theta) = J_1(\theta) J_2(\theta).$$

Since (9 bis) and the preceding parenthetical remark imply that the trigonometric polynomial ML_3 does not vanish identically, it follows from (28) that the zeros θ of neither function J_j have a finite cluster point. This proves that assumption (†) of [3], p. 118, is satisfied.

It follows that, in order to make the three Theorems (i), (iii), (iv) of [3], pp. 118-119, applicable to both of the above systems (23_j), it is sufficient to show that, whether $j = 1$ or 2, the function J_j must change sign at some $\theta = \theta_0{}^j$, while $L_3(\theta_0{}^j) \neq 0$. If the existence of $\theta_0{}^j$ is shown, then the theorems just mentioned imply that

(I_0) *the system* (23_j) *has solution paths reaching to the origin and satisfying* (10), *where* $\theta_0 = \theta_0{}^j$;

(II_0) *to every solution path of* (23_j) *reaching to the origin, there belongs a number* θ_0 *satisfying* (10) *and* $J_j(\theta_0) = 0$;

(III_0) *the limit* θ_0 *in* (10) *satisfies* (11) *whenever* $L_3(\theta_0) \neq 0$.

After the proof of the existence of $\theta_0{}^j$, there will remain to be shown that these results can be transferred from (23_j) to (1). That (I_0) above implies (I) in (*) will be verified at the beginning of Section **10**; that (II_0) and (III_0) imply (II) in (*) will be shown at the end of Section **10** on the basis of some facts to be collected in Section **9**.

In the proof for the existence of a $\theta = \theta_0{}^j$ at which $J_j(\theta)$ changes sign, recourse can be had to the identity

$$(29) \qquad J_1(\theta) = J_2(\theta + \pi),$$

which is clear from (27_j) and (26_j), where $Q^{\frac{1}{2}} \geqq 0$.

As mentioned at the end of Section **6**, there exist values θ_0 satisfying (21) and having the property that $M(\theta)$ vanishes at θ_0 in an odd order. This means that the product $M(\theta)L_3(\theta)$ must change sign at θ_0. Hence the same is true, by (28), of either $J_1(\theta)$ or $J_2(\theta)$. It follows therefore from (29) that each of the functions $J_j(\theta)$ has a zero (θ_0 or $\theta_0 + \pi$) at which it changes sign. Thus a $\theta_0{}^j$ with the desired properties exists and can be identified with θ_0 for one choice of $j = 1, 2$, and with $\theta_0 + \pi$ for the other choice. Hence (I_0), (II_0), (III_0) are applicable to the system (23_j).

9. It is clear from the proofs of Theorems (i)-(iv) in [3,] pp. 119-122, that if θ^0 (in contrast to θ_0) is any angle satisfying $J_1(\theta^0) \neq 0$, then there exists an $s > 0$ having the following property: If

$$(30) \qquad x = x(t), \quad y = y(t), \quad \text{where} \quad t_1 \leqq t \leqq t_2,$$

is any solution of (23_1) and is within the circle C_s: $x^2 + y^2 \equiv r^2 < s^2$ for every t between t_1 and t_2, then, as t increases, the arc (30) can cross the half-line

$$(31) \qquad \theta = \theta^0 \qquad (0 < r < \infty)$$

only in one and the same direction, that is, either with increasing $\theta(t)$ only or with decreasing $\theta(t)$ only, where $\theta(t) = \arctan y(t)/x(t)$. Moreover (cf. *loc. cit.*), if

$$(32) \qquad \theta^1 < \theta < \theta^2 \pmod{2\pi}, \qquad (0 < r < \infty),$$

is a wedge in the (x, y)-plane having the property that the function $J_j(\theta)$ changes sign on the interval $\theta^1 < \theta < \theta^2$, and if (30) is any solution arc of (23_1) which is within the circle C_s and enters the wedge (32) at some t-value, then it cannot leave (32) at a larger t-value. Analogous remarks apply to $J_2(\theta)$ and the solution paths of (23_2).

10. As in Section **2**, let S_1, S_2 denote the two systems of solution paths of (1). It is seen from (4) that, after a suitable numeration $(j = 1, 2)$ of these systems, it can be assumed that if $(x(\tau), y(\tau))$ is a solution path belonging to S_j and if it does not pass through a zero $(x, y) \neq (0, 0)$ of the coefficient $c = c(x, y)$ of (1), that is, if $c(x(\tau), y(\tau)) \neq 0$, then this solution path of (1) can be reparametrized into a solution path $(x(t), y(t))$ of (23_j). Conversely, if $(x(t), y(t))$ is a solution path of (23_j) and satisfies (10) with a θ_0 subject to the restriction $L_3(\theta_0) \neq 0$, then $c(x(t), y(t)) \neq 0$ holds as soon as $(x(t), y(t))$ is close enough to $(0, 0)$. In fact, if $(x, y) \to (0, 0)$ and $c(x, y) = 0$, then $\arctan y/x \to \theta^0$ where $L_3(\theta^0) = 0$. This follows from (2), (3) and the case $k = 3$ of (7) (unless $L_3(\theta) \equiv 0$, a possibility which, in view of Section **6**, can be disregarded). Finally, it is clear that if $(x(t), y(t))$ is a solution path of (23_j) satisfying $c(x(t), y(t)) \neq 0$ throughout, then it is a solution path of (1) contained in the system S_j.

In view of result (I_0) of Section **8** on the solution paths of (23_j), this proves assertion (I) of (*) for the solution paths of (1). It remains to show that assertion (II) of (*) can be deduced from the analogous statements (II_0), (III_0).

Consider a solution path $(x(\tau), y(\tau))$ of (1) which belongs, for example, to S_1 and tends, as $\tau \to \tau_0 - 0$, to the origin $(0, 0)$ in such a way that $c = c(x(\tau), y(\tau))$ becomes 0 for certain τ-values arbitrarily close to τ_0. Let θ^0 be an angle $(\mathrm{mod}\, \pi)$ satisfying $L_3(\theta^0) = 0$ and having the property that the function $J_1(\theta)$ changes sign at some point of the interval $\theta^0 < \theta < \theta^0 + \pi$, say at the point $\theta = \theta_0$. Finally, let ϵ be any (sufficiently small) positive number satisfying $J_1(\theta^0 \pm \epsilon) \neq 0$ and $J_1(\theta^0 + \pi \pm \epsilon) \neq 0$. Then it follows from Section **9** that, when τ is close enough to τ_0, either every or no point of

the path $(x(\tau), y(\tau))$ is within the wedge (32) belonging to $\theta^1 = \theta^0 + \epsilon$, $\theta^2 = \theta^0 + \pi - \epsilon$. But it is clear from (1) and (4) that the first of these two cases is impossible, since $c(x(\tau), y(\tau))$ is supposed to become 0 at certain τ-values arbitrarily close to τ_0. Consequently, as $\tau \to \tau_0$,

(33) $\theta^0 + \epsilon < \theta(\tau) < \theta^0 + \pi - \epsilon$ does not hold for any τ,

where $x(\tau) = r(\tau) \cos \theta(\tau), y(\tau) = r(\tau) \sin \theta(\tau)$; cf. (5)-(6).

Since $\epsilon > 0$ can be chosen arbitrarily small, the boundaries of the wedge prohibited by (33) can be made to be arbitrarily close to the half-lines $\theta = \theta^0 - \epsilon$ and $\theta = \theta^0 + \pi + \epsilon$. Consider, for instance, the former half-line. Then, if the path $(x(\tau), y(\tau))$ crosses at all this half-line (belonging to $\theta^0 - \epsilon$) at a certain $\tau = \tau^0 (< \tau_0)$ close to τ_0, then it follows from Section **9** that the path cannot recross the half-line $\theta = \theta^0 - \epsilon$ at any later $\tau (> \tau^0)$ unless the solution path $(x(\tau), y(\tau))$ contained in S_1 cannot be reparametrized into a solution path $(x(t), y(t))$ of (23_1); that is, unless both functions c, $-b-(b^2-ac)^{\frac{1}{2}}$ of (x, y) vanish at some point of the path $(x(\tau), y(\tau))$, where $\tau^0 < \tau < \tau_0$. Hence, if $(x(\tau), y(\tau))$ crosses the line $\theta = \theta^0 - \epsilon$ at a $\tau = \tau^0$ close to τ_0 in such a way that $\theta(\tau)$ is decreasing, then it cannot recross the line at a later $\tau (> \tau^0)$ before it has crossed the half-line $\theta = \theta^0 + \pi + \epsilon$. In this argument, the half-lines $\theta = \theta^0 - \epsilon$ and $\theta^0 + \pi + \epsilon$ can be interchanged.

It follows that the solution path $(x(\tau), y(\tau))$ of S_1 either satisfies $c(x(\tau), y(\tau)) \neq 0$ for τ near τ_0 (hence is a solution of (23_1)) or one of the following three contingencies must take place:

(34′) $\theta(\tau) \to \theta^0$, (34″) $\theta(\tau) \to \theta^0 + \pi$,

(35) $\theta^0 = \liminf \theta(\tau) < \limsup \theta(\tau) = \theta^0 + \pi$,

where $\tau \to \tau_0$.

In the first case, (II_0) of Section **8** implies that there exists a number θ_0 satisfying (10) and $J_1(\theta_0) = 0$. At this stage of the proof, it is conceivable that the limit θ_0 is the number θ^0 or $\theta^0 + \pi$, so that (34′) or (34″) holds. It will however be shown that, for a solution path $(x(\tau), y(\tau))$ of S_1 reaching to the origin, none of the possibilities (34′), (34″), (35) can hold. The elimination of (35) implies, therefore, that (10) holds. The elimination of (34′) and (34″) implies that the limit θ_0 in (10) does not satisfy $L_3(\theta_0) = 0$. Consequently, (III_0) in Section **8** shows that (11) is a consequence of (10). Hence, the proof of (*) will be complete if it is shown that each of the three contingencies (34′), (34″) and (35) leads to a contradiction.

Ad (34′)-(34″). The angle θ^0, introduced before formula (33) above, was chosen so as to satisfy the condition $L_3(\theta^0) = 0$ and therefore, in view of

(7), the condition $L_3(\theta^0+\pi)=0$ as well, and (34′) or (34″) means that (10) is satisfied (by $\theta_0=\theta^0$ or $\theta_0=\theta^0+\pi$). Let θ_0 be a zero of $M(\theta)$ and let $\vartheta=\theta^0-\theta_0$ or $\vartheta=\theta^0+\pi-\theta_0$ according as (34′) or (34″) holds. Then, after a suitable rotation (19), it can be supposed that (22) holds; thus

$$L_3(\theta_0+\vartheta)\neq 0 \quad\text{and}\quad M(\theta_0+\vartheta)\neq 0.$$

Furthermore, since the limit of $\theta(\tau)$ is invariant under rotations, $\theta(\tau)\to\theta_0+\vartheta$ as $\tau\to\tau_0$. Since $\theta\to\theta_0+\vartheta$ and $L_3(\theta_0+\vartheta)\neq 0$, it follows that $c(x(\tau),y(\tau))\neq 0$ for τ sufficiently near τ_0. Hence $(x(\tau),y(\tau))$ can be reparametrized as a solution of (23_1). But then (II_0) in Section **8** is applicable and claims that $\theta_0+\vartheta$ must be a zero of the function $J_1(\theta)$. It follows therefore from (28) that $M(\theta_0+\vartheta)L_3(\theta_0+\vartheta)=0$. Hence, the last formula line contains a contradiction.

Ad (35). This contingency can be ruled out in the same way as (34′)-(34″) above. In fact, the angular distance (mod 2π) between a zero of $L_3(\theta)$ and a zero of $M(\theta)$ is not, whereas the corresponding distance between the

$$\theta^0=\liminf\theta(\tau)=\liminf\arctan y(\tau)/x(\tau)$$

of (35) and a zero of $M(\theta)$ is, invariant under a rotation (19) of the (x,y)-plane.

11. *Application of* (*) *to lines of curvature.* Let $z=z(x,y)$ be a function of class C^2 in a vicinity of $(x,y)=(0,0)$, and let $K=K(x,y)$ denote the Gaussian, and $H=H(x,y)$ the mean, curvature on the surface $z=z(x,y)$. Then, as is well-known, $H^2\leqq K$, where the sign of equality is characteristic of points (x,y) which are umbilical ("spherical" or "flat" according as $K>0$ or $K=0$, while $K<0$ is precluded by $H^2=K$). It will be assumed that

$$H^2(x,y)\gtreqq K(x,y) \text{ according as } x^2+y^2\gtreqq 0 \tag{36}$$

(which means that $(0,0)$ is an isolated umbilical point) and that

$$K(0,0)\gtreqq 0 \tag{37}$$

(so that $(0,0)$ can be either a spherical or a flat point). Note that (36) and (37) are compatible with the case $K(x,y)<0$, where $(x,y)\neq(0,0)$ (in which case $(0,0)$ must be a flat point), and not only with the usual case $K(x,y)>0$ (in which case $(0,0)$ is a spherical point or a flat point, depending on the alternative in (37)).

If $p=p(x,y),\cdots,t=t(x,y)$ denote the five partial derivatives

$z_x, \cdots, z_{yy}$, the differential equations defining the lines of curvature are defined by the case

$$a = pqr - (1 + p^2)s, \qquad 2b = (1 + q^2)r - (1 + p^2)t, \qquad c = (1 + q^2)s - pqt \tag{38}$$

of (1). Condition (36) means that the resulting quadratic differential equation (1) can or cannot be reduced to non-singular differential equations in a vicinity of a point according as the latter is not or is the point $(0,0)$.

Suppose that $z = z(x, y)$, instead of being just of class C^2, is of class C^3 (as will be seen in a moment, somewhat less would suffice at the point $(0,0)$, without any additional refinement of the C^2-condition at the other points). Then, if the plane tangent to the surface at $(0,0)$ is chosen to be the (x,y)-plane, and if the orientation of the z-axis is suitably chosen, it follows from (36) and (37) that, as $r = (x^2 + y^2)^{\frac{1}{2}} \to 0$,

$$z(x, y) = \tfrac{1}{2}K(0,0)(x^2 + y^2) + \phi(x, y)/6 + o(r^3), \tag{39}$$

where, if α, β, γ, δ denote the partial derivatives of third order of $z(x, y)$ at $(0,0)$, the second term on the right is defined by the cubic form

$$\phi(x, y) = \alpha x^3 + 3\beta x^2 y + 3\gamma x y^2 + \delta y^3. \tag{40}$$

The C^3-assumption also implies that the partial derivatives of first, second and third order, for (x, y) near $(0,0)$, can be obtained by formal differentiations of the Taylor relation (40) (with $o(r^3)$ replaced by $o(r^2)$, $o(r)$ and $o(1)$, respectively). If this is substituted into (40), it is seen that (1) will satisfy conditions (2) and (3) of (*), the values of the six constants α_k, β_k in (2) being those for which the three forms (7) become

$$L_3 = \beta \cos\theta + \gamma \sin\theta = -L_1, \quad L_2 = \tfrac{1}{2}(\alpha - \gamma)\cos\theta - \tfrac{1}{2}(\beta - \delta)\sin\theta. \tag{41}$$

Since the Gaussian and mean curvatures are

$$K = (rt - s^2)/d^4, \text{ where } d = (1 + p^2 + q^2)^{\frac{1}{2}}, \tag{42}$$

and

$$H = I/d^3, \text{ where } I = \tfrac{1}{2}(1 + p^2)t - pqs + \tfrac{1}{2}(1 + q^2)r, \tag{43}$$

it is easily verified from (38) that assumption (4) of Theorem (I) is now equivalent to (36). Hence, in order to render (*) applicable, only its assumptions (7 bis), (8 bis), (9 bis) remain to be assured. But if (41) is inserted into (9), it is seen that (9 bis) is satisfied if and only if

$$\phi(x, y) \not\equiv 0 \tag{44}$$

holds for the cubic form (40). On the other hand, (8) shows that (8 bis) is satisfied if and only if not both linear forms (41) vanish identically, a condition which, in view of (40), is readily found to be equivalent to (44). Finally, condition (7 bis) requires the linear independence of the two linear forms (41), and this condition is satisfied if and only if

$$(45)\qquad \alpha\gamma-\beta\delta\neq(\gamma+\beta)(\gamma-\beta).$$

Hence the situation is as follows:

Let $z=z(x,y)$ be a surface of class C^3 satisfying (36) and suppose that, when the surface is written in the form (39), the associated cubic form (40) satisfies (44) and (45). Then Theorem (*) of Section **3** is applicable to the lines of curvature near the umbilical point $(0,0)$.

Accordingly, every line of curvature which reaches to the umbilical point $(0,0)$ has there a tangent (and the latter is the limit, as $(x,y)\rightarrow(0,0)$, of the tangents at the non-umbilical points (x,y) of that line of curvature). If S_1 and S_2 are the two families of lines of curvature in a vicinity of the umbilical point (cf. the remark which precedes (5) in Section **2**), then both S_1 and S_2 contain at least one curve reaching to $(0,0)$. (Note that the curves contained in either family, say in S_1, are transversal to those contained in S_2, if the umbilical point is excluded.) Finally, if C is a line of curvature which has at $(0,0)$ a continuous tangent and passes *through* the point $(0,0)$ (instead of just reaching it), then *the two arcs into which* $(0,0)$ *divides C cannot be in one and the same family S_j.*

All of this is in agreement with (but is not of course contained in) the particular results derived by Darboux [1], pp. 448-465, as illustrated, in part, by his diagrams, p. 455; cf. also [2], pp. 84-93.

Remark. It may be mentioned that (45) admits of a simple interpretation, as follows: While (36) requires that H^2-K should tend to 0 as $(x,y)\rightarrow(0,0)$, the meaning of (45) is that this limit process should not take place with an exceptional rapidity, but in such a way, for some positive constant,

$$(46)\qquad H^2(x,y)-K(x,y)\geqq \text{const.}\, r^2 \text{ as } r=(x^2+y)^{\frac{1}{2}}\rightarrow 0.$$

In fact, it is seen from (38)-(39) and (42)-(43) that (46) is equivalent to the refinement (13) of (12). But (8) shows that (13) will be satisfied by the two forms (41) if and only if the latter are linearly independent, a restriction which is equivalent to (45).

Incidentally, since the functions (38) are of class C^1 when the surface is of class C^3, the restriction (46) becomes superfluous if use is made of the

C^1-criterion mentioned in Section **5**. In fact, (46) is equivalent to (45), whereas (45) was seen to be equivalent to (7 bis) in the present case.

12. *Application of* (*) *to asymptotic lines.* The assumptions (36)-(37) mean that $(0,0)$ is an isolated umbilical point. If the problem of lines of curvature is replaced by that of the asymptotic lines, the resulting dual situation is as follows: $(0,0)$ is an isolated flat point and the surface is of negative Gaussian curvature near $(0,0)$. This means that

$$(47) \qquad K(x,y) \lesseqq 0 \text{ according as } x^2+y^2 \gtreqq 0$$

and that

$$(48) \qquad H_0 = K_0 = 0, \text{ i. e., } r_0 = s_0 = t_0 = 0$$

$(r = z_{xx}, \cdots)$, where f_0 denotes the value of $f(x,y)$ at $(x,y)=(0,0)$. Clearly, these two conditions mean that, if $(x,y) \neq (0,0)$, there exist on the surface $z = z(x,y)$ (of class C^2) two distinct asymptotic directions, and that the latter become indeterminate (instead of uniting in a single determinate direction) at $(x,y)=(0,0)$.

Let the plane tangent to the surface at $(0,0)$ be chosen to be the (x,y)-plane. Then $z_0 = p_0 = q_0 = 0$, hence the second formulation in (21 bis) shows that $z(x,y) = o(r^2)$ as $(x^2+y^2)^{\frac{1}{2}} = r \to 0$ (Taylor). Hence, if $z(x,y)$ is of class C^3, and if $\phi(x,y)$ denotes the cubic form the coefficients of which are the same as in (39), then

$$(49) \qquad z(x,y) = \phi(x,y) + o(r^3),$$

and the remark made after (39), concerning the formal differentiability of the o-term, holds for (49) also. Since the differential equations defining the asymptotic lines on the surfaces $z = z(x,y)$ result by choosing

$$(50) \qquad a = r, \qquad b = s, \qquad c = t$$

in (1), it follows that, when the surface is of class C^3, conditions (2) and (3) of Theorem (*) are satisfied, the six constants occurring in (2) being given by

$$(51) \qquad (\alpha_1, \beta_1; \alpha_2, \beta_2; \alpha_3, \gamma_3) = (\alpha, \beta; \beta, \gamma; \gamma, \delta),$$

where $\alpha, \beta, \gamma, \delta$ are the coefficients of the cubic (40) occurring in (49). It is also seen from (50) that, in view of (42) and (47), condition (4) is satisfied. Hence, Theorem (*) will be applicable if the four constants occurring on the right of (51) are subject to the conditions required by (7 bis), (8 bis), (9 bis).

First, it follows from (7), (51) and (8) that condition (8 bis) is equivalent to the non-identical vanishing of the Hessian of the cubic form (40), that is, to the restriction

$$(52) \qquad \phi_{xx}\phi_{yy} - \phi_{xy}^2 \not\equiv 0, \text{ i. e., } \phi \not\equiv \lambda^3,$$

where $\lambda = \lambda(x, y)$ denotes an arbitrary linear form (the equivalence of the two assumptions (52) on the cubic form $\phi = \phi(x, y)$ is Hesse's lemma on binary forms ϕ). On the other hand, since (7), (51), (9) and (40) imply that $M(\theta) = \phi(\cos\theta, \sin\theta)$, condition (9 bis) is equivalent to (44) and is therefore implied by the preceding condition (52). Finally, this condition takes care of (7 bis) also. In fact, (40) shows that (52) can be written in the form

$$(\beta^2 - \alpha\gamma)x^2 + (\beta\gamma - \alpha\delta)xy + (\gamma^2 - \beta\delta)y^2 \not\equiv 0,$$

and is therefore equivalent to the non-vanishing of at least one of the determinants

$$\alpha\gamma - \beta^2, \qquad \alpha\delta - \beta\gamma, \qquad \beta\delta - \gamma^2,$$

a condition which, in view of (7) and (51), is identical with the restriction (7 bis).

13. In order to deal with the statements of Section **5**, concerning the case in which

$$(53) \qquad L_1(\theta), L_2(\theta), L_3(\theta) \text{ have a common zero,}$$

it is convenient first to deduce a normalization which belongs to (53) in the same way as the normalization obtained in Section **7** belongs to (7 bis).

Since (53) is the negation of (7 bis), the relations (15)-(18 bis) apply. Let $P(\theta)$ denote the quadratic factor on the right of (18). Then, since $M = PL$, it is seen from (15), (18) and (18 bis) that $M = 0$ must have simple (real) roots only; so that, since L is linear, P must possess at least one simple zero which is not a zero of L. Hence it is seen from the identity (20_3), and from the corresponding identity

$$(54) \qquad L^*(\theta) = L(\theta - \theta^*),$$

that $M^*(\theta)$ possess at least one simple zero which is not a zero of $L^*(\theta)$.

It is readily verified from (20_1) and (16) that $L_1^*(\theta)$, $L_2^*(\theta)$, $L_3^*(\theta)$ are respectively identical with $P(\theta^*)$, $\frac{1}{2}dP(\theta^*)/d\theta^*$, $P(\theta^* + \frac{1}{2}\pi)$ times $L(\theta + \theta^*)$. This, when combined with the preceding remarks, implies that, if the constant θ^* occurring in (19) is suitably chosen, then none of the

three functions $L_k^*(\theta+\theta^*)$ of θ will vanish identically and the product $L_1^*(\theta)L_3^*(\theta)$ will be non-positive throughout. (Note that if the zeros of the quadratic trigonometric polynomial $P(\theta)$ differ from each other by multiples of $\frac{1}{2}\pi$ only, then $L_1^*(\theta)L_3^*(\theta)$ cannot become non-negative unless $L_3^*(\theta)$ vanishes identically.)

Let the constant θ^* defining the rotation (19) be suitably chosen and then all asterisks omitted (so that $\theta^*=0$). Then the above remarks can be summarized as follows:

If (53) holds instead of (7 bis), then there is no loss of generality in assuming that

(55) $$L_k(\theta)\not\equiv 0, \text{ where } k=1,2,3,$$

that

(56) $$L_1(\theta)L_3(\theta)\leqq 0 \text{ for all } \theta,$$

finally that there exists a θ_0 satisfying

(57) $$L_3(\theta_0)\neq 0 \text{ and } M(\theta_0)=0 \text{ but } M_\theta(\theta_0)\neq 0,$$

where $M_\theta(\theta)=dM(\theta)/d\theta$.

14. It is now easy to prove the statement made at the end of Section **5.** The statement deals with the case in which (7 bis) is replaced by the assumption that all (real) roots of (9) are simple (i. e., that

(58) $$M_\theta(\theta)\neq 0 \text{ whenever } M(\theta)=0,$$

where $M_\theta=dM/d\theta$), and runs as follows:

(†) *If assumption* (7 bis) *of* (*) *is replaced by* (58), *then the assertions of* (*) *are true at least in the sense of* (§), *Section* **5.**

Proof. Since (57) means that $M(\theta)$ changes signs at $\theta=\theta_0$, while $L(\theta_0)\neq 0$, a major part of the proof of (*) remains applicable. In fact, the proof of assertion (I) remains valid, as does that portion of the proof of (II) according to which either the limit (10) exists or (35) holds, where $L(\theta^0)=0$ (the first of these alternative cases includes the contingencies (34′), (34″) in the present situation). Also, (10) implies (11) if the limit θ_0 in (10) is not a zero of $L(\theta)$. Thus it only remains to show that (53) and (50) exclude the possibility of (35).

According to the normalization (55), no c_k can vanish in (15). Since the equations (1), (2), $\cdots$, being homogeneous, can be multiplied by an arbitrary non-vanishing constant, it follows that it can be supposed that $c_3=1$,

that is, that $L_3(\theta) = L(\theta)$. Then (25) reduces to

$$F(\theta) = L(\theta) \tag{59}$$

and (26) to

$$G_j(\theta) = L(\theta)(-c_2 + (-1)^j (c_2{}^2 - c_1c_3)^{\frac{1}{2}} \operatorname{sgn} L(\theta)), \tag{60}$$

by (17). Put

$$N_j(\theta) = (-c_2 + (-1)^j (c_2{}^2 - c_1c_3)^{\frac{1}{2}}) \cos\theta - c_3 \sin\theta. \tag{61}$$

Then the definition (27_j) of $J_j(\theta)$ shows that

$$J_j(\theta) = L(\theta) N_i(\theta) \tag{62}$$

holds in the half-plane $L(\theta) \geqq 0$ or $L(\theta) \leqq 0$ according as $j = i$ or $j \neq i$. Since $c_3 = 1$, it is readily verified that $P = N_1N_2$, hence $M = LN_1N_2$. It follows therefore from (58) that the zeros of N_j are distinct from those of L. Consequently, (62) shows that $J_j(\theta)$ changes sign in both of the half-planes

$$\theta^0 < \theta < \theta^0 + \pi, \qquad \theta^0 + \pi < \theta < \theta^0 + 2\pi$$

if $L(\theta^0) = 0$.

Accordingly, if $(x(t), y(t))$ is a solution path of (23_j) on some t-interval, is within the circle $x^2 + y^2 < s^2$ on this interval (for a sufficiently small s) and enters one of the wedges

$$\theta^0 + \epsilon < \theta < \theta^0 + \pi - \epsilon, \qquad \theta^0 + \pi + \epsilon < \theta < \theta^0 + 2\pi - \epsilon,$$

then it cannot leave that wedge; cf. the remarks concerning (32). This fact eliminates the possibility of (35), since a solution path of (1) in such a wedge is a solution path of (23_j) (either for $j = 1$ or for $j = 2$), and conversely.

15. There will now be proved the C^1-criterion announced in Section **5**, that is, the following theorem:

(**) *If assumption* (7 bis) *of* (*) *is omitted but the coefficient functions a, b, c of* (1) *(which in* (*) *are required to be just continuous) are assumed to have continuous partial derivatives $a_x, \cdots, c_y$, then the assertions of* (*) *remain true at least in the sense of* (§), Section **5**.

Proof. In view of (*) and (†), it will be sufficient to prove this theorem only for the case excluded by (*) and (†) together, that is, for the case in which neither the assumption (7 bis) of (*) nor assumption (58) of (†) is

fulfilled. Since there exists a $\theta = \theta_0$ satisfying (57) in this case also, the proofs of (*) and (†) show that (**) will be proved if contingency (35) is eliminated for every solution path of (1) reaching to the origin.

The considerations of Section **13** make it clear that there is no loss of generality in assuming that $\theta^0 \neq \frac{1}{2}\pi \pmod{\pi}$ if $L(\theta^0) = 0$. This means that the coefficient β of $\sin\theta$ in (15) is not 0. Since $f_k(x, y)$ in (2)-(3) is supposed to be of class C^1, it follows that the partial derivative of a, b, c with respect to y at $(0, 0)$ is $c_1\beta$, $c_2\beta$, $c_3\beta$, respectively. Since $c_k \neq 0$ and $\beta \neq 0$, continuity considerations show that a_y, b_y, c_y are distinct from 0 for (x, y) near $(0, 0)$. Hence, the k-th of the equations

$$(63_1)\quad a(x,y) = 0; \qquad (63_2)\quad b(x,y) = 0; \qquad (63_3)\quad c(x,y) = 0$$

has a unique solution $y = y(x)$ of class C^1 for small $|x|$, say the solution

$$(64_k)\qquad y = y_k(x), \text{ where } y_k(0) = 0 \text{ and } dy_k(0)/dx = \tan\theta^0.$$

Since (4) implies that a point $(x, y) \neq (0, 0)$ of the curve (63_2) cannot be on either of the curves (63_1), (63_3), and since (63_k) is equivalent to (64_k) for every fixed k, it is clear that

$$(65)\qquad \text{either } y_2(x) > y_1(x) \text{ or } y_2(x) < y_1(x) \text{ for all small } x > 0$$

and that

$$(66)\qquad \text{either } y_2(x) > y_3(x) \text{ or } y_2(x) < y_3(x) \text{ for all small } x > 0,$$

finally that such alternatives hold for all small $x < 0$ also.

On a τ-interval $0 \leqq \tau < \tau_0$, let

$$(67)\qquad x = x(\tau), \qquad y = y(\tau)$$

be a solution path of (1) satisfying (5) as $\tau \to \tau_0$, and let (67) belong, for example, to the family S_1. On a sufficiently small τ-vicinity of any fixed τ, the coordinates of the solution path (67) must satisfy at least one of the differential equations

$$(68)\qquad dy/dx = \{-b - (b^2 - ac)^{\frac{1}{2}}\}/c, \qquad dx/dy = c/\{-b - (b^2 - ac)^{\frac{1}{2}}\}$$

(where $\{\quad\}/c$ is meant to represent $-\frac{1}{2}a/b$ if $c = 0$ and $b < 0$, hence $-b - (b^2 - ac)^{\frac{1}{2}} = 0$), as well as at least one of the differential equations

$$(69)\qquad dx/dy = \{-b + (b^2 - ac)^{\frac{1}{2}}\}/a, \qquad dy/dx = a/\{-b + (b^2 - ac)^{\frac{1}{2}}\}$$

(with an analogous interpretation of $\{\quad\}/a$ if $a = 0$ and $b > 0$). It is clear from these differential equations that, at a given τ, the first of the

function (67) has a relative extremum (maximum or minimum) if and only if

(70) $c(x(\tau), y(\tau))$ changes signs and $b(x(\tau), y(\tau)) > 0$

at that τ, and that the second of the functions (67) has a relative extremum if and only if

(71) $a(x(\tau), y(\tau))$ changes signs and $b(x(\tau), y(\tau)) < 0$.

On the basis of these facts, the possibility of (35) can be ruled out as follows:

Suppose, if possible, that (35) is satisfied by the solution path (67) of (1) reaching to the origin, as $\tau \rightarrow \tau_0 - 0$. Then, corresponding to every $\epsilon > 0$, and for every τ close enough to τ_0, the path (67) is in the wedge

$$\theta^0 - \epsilon < \theta < \theta^0 + \pi + \epsilon \qquad (r > 0), \tag{72}$$

and (67) crosses the half-line

$$\theta = \theta^0 + \epsilon \qquad (r > 0) \tag{73}$$

an infinity of times as $\tau \rightarrow \tau_0$. On the half-line (73) (for small r), the slope dy/dx in (68) and/or (69) is of constant sign. It follows that each of the functions (67) has an infinity of extrema as $\tau \rightarrow \tau_0$. These extremal values must be attained when $(x(\tau), y(\tau))$ is in one of the wedges $|\theta - \theta^0| < \epsilon$, $|\theta - \theta^0 - \pi| < \epsilon$, since these wedges contain the arcs (64_k). This follows from the above criteria involving (70) and (71).

The normalization (55) and (2)-(3) imply that, if s is sufficiently small, the sector

$$\theta^0 - \epsilon < \theta < \theta^0 + \epsilon, \qquad 0 < r < s \tag{74}$$

is divided by the arc (64_k) into two domains on which $a_k > 0$ and $a_k < 0$, respectively, where $a_1 = a$, $a_2 = b$, $a_3 = c$. Since both (70) and (71) occur infinitely often as $\tau \rightarrow \tau_0$, it follows that the arc (63_1) is in the set $b < 0$ and that the arc (63_3) is in the set $b > 0$; cf. (65) and (66). Hence, exactly one of the functions a, c has opposite signs on the half-line (73) and on the arc (63_2). (For example, if $b > 0$ on (73) for small $r > 0$, then c has opposite signs on (73) and on $b = 0$, while a is of the same sign.)

According to (4), $a(x, y)c(x, y) < 0$ if $b(x, y) = 0$ and $(x, y) \neq (0, 0)$. On the other hand, the normalization (56) implies that

$$L_1(\theta^0 + \epsilon)L_3(\theta^0 + \epsilon) < 0;$$

hence, by (2)-(3), $a(x, y)c(x, y) < 0$ on the half-line (73) for small $r > 0$. Clearly, this contradicts the fact that exactly one of the functions a, c has opposite signs on (63_2) and on (73).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 4 (1896).

[2] A. Gullstrand, "Zur Kenntniss der Kreispunkte," *Acta Mathematica*, vol. 29 (1905), pp. 59-100.

[3] P. Hartman and A. Wintner, "On the behavior of the solutions of binary differential systems at singular points," *American Journal of Mathematics*, vol. 75 (1953), pp. 117-126.

[4] H. Liebmann, "Geometrische Theorie der Differentialgleichungen," *Encyklopädie der Mathematischen Wissenschaften*, article III D8 (1914); more particularly, pp. 516-517.

[5] E. Picard, *Traité d'Analyse*, vol. 3, 3rd ed., 1928. (In all editions, the relevant section is a reproduction of [6] below.)

[6] ———, "Sur les points singuliers des équations différentielles du premier ordre," *Mathematische Annalen*, vol. 46 (1895), pp. 521-528.

[7] A. Wahlgren, "Sur les points singuliers des équations différentielles du premier ordre et du second degré," *Bihang till K. Svenska Vet.-Akad. Handlingar*, vol. 28 (1902), no. 4.

[8] A. Wintner, "Unrestricted Riccatian solution fields," *Quarterly Journal of Mathematics* (Oxford), vol. 18 (1947), pp. 65-71.

ON THE THIRD FUNDAMENTAL FORM OF A SURFACE.*

By PHILIP HARTMAN and AUREL WINTNER.

Part I.

1. Let D be an open, simply connected (and, for the purposes at hand, sufficiently small) domain in a (u,v)-plane, and let $g_{11}, g_{12} = g_{21}, g_{22}$ be three functions on D corresponding to which the quadratic form

$$(1) \qquad g_{\alpha\beta}(u^1, u^2)\, du^\alpha du^\beta, \qquad \text{where } u^1 = u,\, u^2 = v,$$

is positive definite at every point of D. Suppose further that the three functions $g_{ik}(u,v)$ are of class C' (i. e., that the partial derivatives Γ_u, Γ_v of the "vector"

$$(2) \qquad \Gamma = (g_{11}, g_{12}, g_{22})$$

exist and are continuous on D). Then (1) will be referred to as a *C'-metric (on D)*. Since the second derivatives of (2) need not exist, (1) will not in general have a Gaussian curvature $K = K(u,v)$. On the other hand, since

$$(3) \qquad g \neq 0, \quad \text{where } g = (\det g_{ik})^{\frac{1}{2}},$$

there exist on D continuous Christoffel coefficients

$$(4) \qquad \Gamma^j_{ik} = \Gamma^j_{ik}(u,v) = \Gamma^j_{ki}.$$

This implies that Levi-Civita's parallel transport is uniquely defined along every oriented C'-arc (contained in D) and that there exists a continuous geodesic curvature $\kappa = \kappa(s)$ on every C''-arc.

If

$$(5) \qquad u^* = u^*(u,v), \qquad v^* = v^*(u,v)$$

is a pair of functions which is of class C' and of non-vanishing Jacobian on D, and if D^* denotes the (u,v)-domain on which the (sufficiently small) (u,v)-domain D is mapped by (5), finally if

$$(6) \qquad g^*_{\alpha\beta}(u^{*1}, u^{*2})\, du^{*\alpha} du^{*\beta}, \quad \text{where } u^{*1} = u^*,\, u^{*2} = v^*,$$

* Received October 13, 1952.

is the positive definite form which is identical with the C'-metric (1) by virtue of (5), then (6) need not be a C'-metric on D^*. In fact, if the transformation (5) and its inverse are just of class C', then the functions $g^*_{ik}(u^*, v^*)$, instead of being of class C', will be just continuous on D^*. All that follows is that if (1) is a C'-metric on D, then (6) will be a C'-metric on D^* when the transformation (5) is of class C'' and of non-vanishing Jacobian. Conversely, it was shown in [5] that the transformation (5) must be of class C'' if it is of class C', has a non-vanishing Jacobian and transforms a C'-metric (1) on D into a C'-metric (6) on D^*.

As mentioned above, the Gaussian curvature $K = K(u, v)$ of a C'-metric does not in general exist. It is possible however to define with reference to every C'-metric *the total curvature*

$$\tau = \tau(E) \tag{7}$$

of certain subsets E of D, as follows: If $E = E(J)$ is the interior of an oriented Jordan curve J contained in D and consisting of a finite number of arcs each of which is of class C', let the total curvature (7) of E be defined as the oriented variation (with respect to the tangent vector) of the direction of a vector transported parallel to itself (Levi-Civita) along $J = J(E)$. Clearly, this definition of the set function τ remains invariant under one-to-one C''-transformations (5) which transform the given C'-metric (1) into a C'-metric (6).

If the continuous functions (4), occurring in the definition of (7), are expressed in terms of the partial derivatives $g_{ik\,u}$, $g_{ik\,v}$, it is readily found that

$$\tau(E) = \int_J (2g_{11}g)^{-1}\{(g_{12}g_{11\,u} + g_{11}g_{11\,v} - 2g_{11}g_{12\,u})du + (g_{12}g_{11\,v} - g_{11}g_{22\,u})dv\} \tag{8}$$

(cf. [1], pp. 123; this corrects the explicit form of the integrands in formulae (3)-(6) of [10], pp. 877-878). A partial integration (Green) transforms the explicit representation (8) of (7) into

$$\tau(E) = -\iint_E (4g^3)^{-1}\det(\Gamma, \Gamma_u, \Gamma_v)\,du\,dv + \int_J (2g)^{-1}\{(g_{11\,v} - g_{12\,u})du + (g_{12\,v} - g_{22\,u})dv\}, \tag{9}$$

where g^3 denotes the cube of (3) and Γ_u, Γ_v are the partial derivatives of (2).

Without using these explicit representations of (7), the following definition will now be introduced: Let a C'-metric (1) on D be called a *Gaussian metric* or such as to possess a *curvature* $K = K(u, v)$ if the set function (7), representing the total curvature of an arbitrary $E = E(J)$, is absolutely continuous, that is, if there exists on D a point function $K = K(u, v)$ which is integrable (L) on every compact subset of D and is such that, if $g = g(u, v)$ is defined by (3), then

$$\tau(E) = \iint_E K g\, du\, dv \tag{10}$$

holds for every $E = E(J)$. This definition of the curvature, initiated by Weyl ([18], pp. 42-44; cf. [13], p. 135) leaves K (if it exists) undetermined on sets of measure 0. It is however clear what will be meant by a C'-metric which possesses a bounded curvature $K(u, v)$, a continuous curvature $K(u, v)$, etc.

Under the assumption that the C'-metric (1) is *Gaussian*, let (8*), (9*) denote the relations which result if (10) is substituted into (8), (9), respectively. Then the above introduction of a *curvature* K is justified by the following pair of facts (neither of which is contained in the other): On the one hand, if the coefficient functions of (1) are of class C'', so that the classical formula for the Gaussian curvature defines a (continuous) function $K = K(u, v)$, and if the latter is substituted into (10), then the resulting set function (10) satisfies (8*). On the other hand, if (1) is the C'-metric on a surface $X = X(u, v)$ of class C'', where $X = (x, y, z)$, and if the (continuous) function $K = K(u, v)$ is defined to be the quotient of the determinants of the second and first fundamental forms of this embedding of (1), then the resulting set function (10) satisfies (9*). The first of these two facts follows by observing that (8*) is an integrated form of Liouville's representation of the Gaussian curvature of a C''-metric (cf. [1], p. 123), while the second fact follows from the circumstance that the embedded form of (9*) is identical with formula (7), p. 759, of [6], a formula which holds on surfaces $X = X(u, v)$ of class C''; cf. [18], pp. 42-44 and [13], p. 135 or [6], p. 760.

2. In order to deal with questions involving the normal image on a sphere, or with the third fundamental form, of a surface in a manner which avoids the usual unnatural restrictions of differentiability (restrictions which lack a direct geometrical significance), the case $K_0 = 1$ of the following theorem will be needed:

(i) *If a C'-metric* (1) *on D possesses a constant curvature* (*in the sense of* (10), *that is, if the set function* (7), *where $E = E(J)$, is representable in the form*

$$(11) \qquad \tau(E) = K_0 \iint_E (\det g_{ik})^{\frac{1}{2}} du dv,$$

where $K_0 = \text{const.} \gtreqless 0$), *then there exists, near every point* (u, v) *of D, a transformation* (5) *which is of class C'' and of non-vanishing Jacobian, and which transforms* (1) *into the standard analytic form*

$$(12) \qquad (du^{*2} + dv^{*2})/f^2, \quad \text{where } f = 1 + (u^{*2} + v^{*2})K_0/4.$$

This theorem (i), which is a refinement (for the two-dimensional case) of the characterization of metrics of constant curvature due to Weingarten (cf. [2]), is contained between the lines of [19]. In fact, the situation is as follows: In order to prove the theorem in its preceding formulation, it is sufficient to show that, if the total curvature (7) of a C'-metric satisfies (11), then Weingarten's system of three linear differential equations

$$(13) \qquad \phi_{ik} - \Gamma^n{}_{ik}(u^1, u^2)\phi_n + K_0 g_{ik}(u^1, u^2)\phi = 0$$

(in which the functions (4) are just continuous and the subscripts of ϕ denote partial differentiations with respect to $u^1 = u$ and $u^2 = v$) is "total." By this is meant that (13) possesses a (unique) solution $\phi = \phi(u, v)$ for which the function ϕ and its first partial derivatives ϕ_1, ϕ_2 reduce to arbitrarily given values, ϕ^0 and $\phi_1{}^0$, $\phi_2{}^0$, at an arbitrary point (u^0, v^0) of D. For, if this is assured, then the above theorem (i) follows by the arguments used in [19], Sections 4 and 7. But the system (13) of three equations of second order will be "total" if it is "total" when written in the form of six equations of first order,

$$(14) \qquad \partial\phi/\partial u^i = \psi_i, \qquad \partial\psi_i/\partial u^k = \Gamma^n{}_{ik}\psi_n - K_0 g_{ik}\phi.$$

In view of theorem (II) in [6], this requires that the system (14) should satisfy the set of the integrability conditions to which the last two formula lines of theorem (II) in [6], p. 761, reduce in the present case, represented by (14). Since (14) is a homogeneous system in three unknown functions ϕ, ψ_1, ψ_2 and two independent variables, there are 9 ($= 3^2$) such integrability conditions. A direct calculation shows that one of these reduces to the identity $0 = 0$; two are satisfied by virtue of $g_{ik} = g_{ki}$ and $\Gamma^j{}_{ik} = \Gamma^j{}_{ki}$; two are equivalent to the conditions

$$\int_J K_0 g_{ik} du^k = \int\int_E K_0 (\partial g_{i2}/\partial u^1 - \partial g_{i1}/\partial u^2) du^1 du^2, \qquad \text{where } i = 1, 2,$$

which are satisfied if and only if K_0 is a constant; and the last four are represented by the equations

$$\int_J \Gamma^j_{in} du^n = \int\int_E (\Gamma^n_{i1}\Gamma^j_{n2} - \Gamma^n_{i2}\Gamma^j_{n1} + (-1)^i g^{ij} K_0 g^2) du^1 du^2,$$

where $i, j = 1, 2$ and $(g^{ik}) = (g_{ik})^{-1}$. These four relations are equivalent, by virtue of the Lemma in [6], p. 761, to four others in which the line integral on the left is replaced by $\int_J g^{-1} g_{jk} \Gamma^j_{in} du^n$, where $i, k = 1, 2$. Among these last four integral conditions, two are trivially satisfied and two reduce to (9), provided that the C'-metric (1) has the curvature K_0. Since this is precisely the assumption of (i), the proof of (i) is complete.

The proof of (i) has the following consequence:

(i bis) *Let* (1) *be a C'-metric on a simply connected domain D; Γ^j_{ik} the Christoffel symbols of the second kind belonging to* (1); *finally, K_0 a continuous function on D. For every point (u_0, v_0) of D and every set of three numbers ϕ^0, ϕ_1^0, ϕ_2^0, there exists a solution $\phi = \phi(u, v)$ of class C'' on D of* (13) *satisfying $\phi(u_0, v_0) = \phi^0$, $\phi_1(u_0, v_0) = \phi_1^0$, $\phi_2(u_0, v_0) = \phi_2^0$ if and only if K_0 is a constant and the metric* (1) *has a curvature K, which is the constant $K = K_0$.*

The proof of (II) in [6], pp. 763-765, on which the proof of (i), (i bis) is based, shows that if Γ^j_{ik}, g_{ik} are arbitrary continuous functions in (13), then (whether or not (13) is "total") (13) has at most one solution $\phi(u, v)$ of class C'' satisfying given initial conditions $\phi(u_0, v_0) = \phi^0$, $\phi_1(u_0, v_0) = \phi_1^0$, $\phi_2(u_0, v_0) = \phi_2^0$.

Part II.

3. A set S of points in the euclidean space $X = (x, y, z)$ will be called a surface of class C^n, where $n \geqq 1$, if there exist some (u, v)-domain D and some vector function $X(u, v)$ of class C^n on D such that the vector product $[X_1, X_2]$, where $X_1 = \partial X/\partial u$, $X_2 = \partial X/\partial v$, does not vanish and $X = X(u, v)$ is a one-to-one mapping of D onto S. The vector function $X = X(u, v)$ will be said to be a C^n-parametrization of S. The classes C^1, C^2, C^3 will be denoted by C', C'', C'''.

Suppose that S is of class C'', and let $X = X(u, v)$ be a C''-para-

metrization of S. Then the unit normal vector

(15) $$N = [X_1, X_2]/|[X_1, X_2]| \qquad ([X_1, X_2] \neq 0)$$

is a function $N(u, v)$ of class C' on D. If the binary symmetric matrices

(16) $\alpha = (a_{ik})$, (17) $\beta = (b_{ik})$, (18) $\gamma = (c_{ik})$

are defined by

(19) $a_{ik} = X_i \cdot X_k$, (20) $b_{ik} = - N_i \cdot X_k$, (21) $c_{ik} = N_i \cdot N_k$,

then $\alpha = \alpha(u, v)$ is of class C', while $\beta = \beta(u, v)$ and $\gamma = \gamma(u, v)$ are continuous (on D) and

(22) $|dX|^2 = a_{ik} du^i du^k$, (23) $-dX \cdot dN = b_{ik} du^i du^k$,

(24) $|dN|^2 = c_{ik} du^i du^k$

are, respectively, the first, second, third fundamental forms on

$$S: X = X(u^1, u^2),$$

where $u^1 = u$, $u^2 = v$.

For reasons which will become obvious in a moment, the coefficients of the second fundamental form are defined by (20), and not by the formula $b_{ik} = N \cdot X_{ik}$, involving the second derivatives of X. However, the possibility of so defining them (for the case at hand) shows that β is a symmetric matrix.

The Gaussian and mean curvatures, K and H, are defined by

(25) $K = \det(\beta\alpha^{-1})$, (26) $H = \frac{1}{2}\mathrm{tr}(\beta\alpha^{-1})$

and are continuous functions on D. The reciprocal α^{-1} of (16) exists, since the parenthetical assumption of (15) means that

(26) (22) is positive definite.

On the other hand,

(27) (24) is positive definite at non-parabolic points only,

that is, if and only if $K \neq 0$. In fact, (27) follows from (24), (22) and from the identity

(28) $$[N_1, N_2] = K[X_1, X_2], \qquad ([X_1, X_2] \neq 0),$$

while (28) follows from (25) and (19)-(20) if use is made of Weingarten's derivation formulae

(29) $$N_j = - b_{ji} a^{ik} X_k, \quad \text{where } j = 1, 2 \text{ and } (a^{ik}) = \alpha^{-1}.$$

Both (28) and (29) hold also at parabolic points (u, v), points at which (24) is positive semi-definite. Even at the latter points, the matrix of (24) is uniquely determined by the matrices of (22) and (23), since

$$\gamma = \beta\alpha^{-1}\beta. \tag{30}$$

In fact, (30) follows if (29) is substituted into (21) and then use is made of (19) and (20); cf. [21], p. 372.

The classical representation of the third fundamental form in the terms of the first and the second is not (30) but

$$\gamma = -K\alpha + 2H\beta \tag{31}$$

with (25)-(26). The standard proof of (31) is open to objections, partly because it excludes umbilical points (points (u, v) at which $H^2 = K$) and which can form a complicated (u, v)-set even if S is of class C^∞, but mainly because it is valid only if S has a C''-parametrization $X = X(u, v)$ in which $u = \text{const.}$ and $v = \text{Const.}$ are lines of curvature, whereas all that is assumed now is that S has *some* C''-parametrization. In order to verify (31) under the latter assumption alone, note that, if ϵ denotes the unit matrix, then, since the definitions (25)-(26) reduce the characteristic equation $\det(s\epsilon - \alpha^{-1}\beta) = 0$ of $\alpha^{-1}\beta$ to $s^2 - 2Hs + K = 0$, the latter equation must be satisfied by $s = \alpha^{-1}\beta$ (Hamilton-Cayley). If the resulting matrix relation is multiplied by α from the right, it follows that (31) is equivalent to (30).

Incidentally, (30) and (25) imply (27) and also show that, at every point (u, v) of D, any of the three assumptions $K = 0$, $\det\beta = 0$, $\det\gamma = 0$ is equivalent to the other two. This implies that the first fundamental form is uniquely determined by the second and the third, in any (u, v)-domain not containing parabolic points, since (30) can be written in the form

$$\alpha = \beta\gamma^{-1}\beta \text{ if (and only if) } K \neq 0. \tag{32}$$

It also follows that, under the assumption of (32), the relations (25), (26) can be written as

$$0 \neq K = \det(\gamma\beta^{-1}), \tag{33}$$ $$H = \tfrac{1}{2}\operatorname{tr}(\gamma\beta^{-1}) \tag{34}$$

(Weingarten).

4. Let a sufficiently small (u, v)-domain D be mapped on a (u^*, v^*)-domain D^* by a transformation (5) of class C' and of non-vanishing Jacobian. Then, if $X = X(u, v)$ is a C''-parametrization (on D) of a surface S of class C'', the parametrization

$$(34) \qquad X = X(u^*; v^*) \equiv X(u(u^*, v^*), v(u^*, v^*))$$

of S (on D^*) will in general be of class C' only. Nevertheless, *the unit normal vector*

$$(35) \qquad N = N(u^*; v^*),$$

defined by that analogue of (15) *in which X_i is replaced by $\partial X/\partial u^{*i}$, is a function of class C' on D^* (instead of being just continuous).* For, on the one hand, $N(u^*; v^*) = \pm N(u(u^*, v^*), v(u^*, v^*))$, where $\pm = \operatorname{sgn} \partial(u, v)/\partial(u^*, v^*)$ and, on the other hand, both $N(u, v)$ and $u = u(u^*, v^*)$, $v = v(u^* v^*)$ are of class C'.

Thus those analogues of (20), (21), (22) in which X_i, N_k are replaced by $\partial X/\partial u^{*i}$, $\partial N/\partial u^{*k}$ lead to continuous matrix functions

$$(36) \quad \alpha^* = (a^*_{ik}), \qquad (37) \quad \beta^* = (b^*_{ik}), \qquad (38) \quad \gamma^* = (c^*_{ik})$$

of (u^*, v^*) on D^*. It is clear that α^* and γ^* are symmetric matrices. In view of $X(u, v) = X(u^*; v^*)$ and $N(u, v) = N(u^*; v^*) \operatorname{sgn} \partial(u, v)/\partial(u^*, v^*)$, the transformation rules $\alpha \to \alpha^*$, $\beta \to \beta^*$, $\gamma \to \gamma^*$ are identical with the standard transformation rules when (5) is a substitution of class C''; more precisely,

$$(39_\alpha) \qquad \alpha^* = \phi'\alpha\phi, \qquad\qquad (39_\gamma) \qquad \gamma = \phi'\gamma\phi,$$

though what would correspond to (39_β) must be replaced by

$$(40) \qquad \beta^* = \pm\, \phi'\beta\phi, \text{ with } \pm = \operatorname{sgn} \det \phi,$$

where ϕ is the Jacobian matrix

$$(41) \qquad \phi = (\partial u^i/\partial u^{*k})$$

and ϕ' is the transpose of ϕ. The symmetry of β and the transformation rule (40) show that β^* is symmetric. In view of the definitions of α^*, β^*, γ^*, the analogues of the relations (25)-(34), obtained by replacing α, β, γ, X_i, N_k by α^*, β^*, γ^*, $\partial X/u^{*i}$, $\partial N/\partial u^{*k}$, remain valid.

5. Let a parametrization $X = X(u^*, v^*)$ of a surface S of class C'' be called a *normal C'-parametrization* if $X = X(u^*, v^*)$ is of class C' (in contrast to certain other parametrizations $X = X(u, v)$ of S, in which $X(u, v)$ is of class C'') and if the normal $N = N(u, v)$ gives a C''-parametrization of a portion of the unit sphere $|N| = 1$. A normal C'-parametrization $X = X(\lambda, \mu)$ of S will be called a *spherical C'-parametrization* if the parameters λ, μ satisfy $\lambda^2 + \mu^2 < 1$ and are two of the three direction cosines of N; for example,

$$(42) \qquad N = (\lambda, \mu, \nu)$$

where

$$(43) \quad \nu = (1 - \rho^2)^{\frac{1}{2}}, \qquad (44) \quad \rho = (\lambda^2 + \mu^2)^{\frac{1}{2}}$$

(or $N = (\nu, \lambda, \mu)$ or $N = (\lambda, \nu, \mu)$).

(ii) *A surface S of class C'' has a normal (and/or spherical) C'-parametrization $X = X(u^*, v^*)$ if and only if S is free of parabolic point, that is,*

$$(45) \qquad K \neq 0 \qquad \text{(i. e., } \det \beta \neq 0 \text{ and/or } \det \gamma \neq 0\text{)}.$$

(ii bis) *A spherical parametrization $X = X(u^*, v^*)$ of S is of class C'' if and only if S is of class C'''* (that is, if and only if S has, besides a C''-parametrization $X = X(u, v)$ which is assumed, some C''-parametrization $X = X(u', v')$ also).

It is understood that S is always meant to be "sufficiently small," that is, that the assertions are local, in the sense of referring to some vicinity of a given point of S.

In order to prove (ii), let $X = X(u, v)$ be a C''-parametrization of S. Then it is clear from (28) that $N = N(u, v)$ is a C'-parametrization of (part of) the sphere $|N| = 1$ if and only if $K \neq 0$ on S. If $K \neq 0$, hence $[N_1, N_2] \neq 0$ in view of (15), then there exists a transformation $(u, v) \rightarrow (u^*, v^*)$, which is of class C' and of non-vanishing Jacobian, and in which (u^*, v^*) is (λ, μ) or (μ, ν) or (λ, ν), finally, which transforms $N(u, v)$ into (42)-(43). Clearly, $X = X(u^*, v^*)$ is a normal C'-parametrization; in fact, it is a spherical C'-parametrization of S. Hence, $K \neq 0$ is sufficient for the existence of spherical C'-parametrizations. Incidentally, this argument makes it clear that spherical parametrizations of a given class $C', C'', \cdots$ exist if and only if normal parametrizations of the same class do.

The necessity of $K \neq 0$ is clear from the definition of normal parametrizations, in which it is required that $N(u^*, v^*)$ be a C''-parametrization of a portion of the unit sphere; in particular, that the vector product of $\partial N/\partial u^*$ and $\partial N/\partial v^*$ be distinct from 0. Since the analogue of (28) holds in C'-parametrizations $X = X(u^*, v^*)$ of a surface of class C', it follows that $K \neq 0$ when S has a normal C'-parametrization. This proves (ii).

The proof of the "if" part of (ii bis) is clear from the proof of the first part of (ii). For if $X = X(u, v)$ is a C'''-parametrization of S, then $N = N(u, v)$ is a C''-parametrization of part of the sphere $|N| = 1$ and the above-described transformation $(u, v) \rightarrow (u^*, v^*)$ is of class C'' and leads to a spherical C''-parametrization. Since a spherical parametrization is

derived from another (when two exist) by an analytic transformation of the parameters, all (one or three) spherical parametrizations are of class C'' when S is of class C'''. The converse, that is, the "only if" part of (ii bis), is contained in the following lemma as a particular case, $u^* = u$, $v^* = v$ and $n = 2$.

LEMMA. *If a surface S, of class C^n for a fixed $n \geqq 1$, possesses a C^n-parametrization $X = X(u, v)$ in which the normal $N = N(u, v)$ becomes a function of class C^n, then S is a surface of class C^{n+1}* (that is, S has some C^{n+1}-parametrization $X = X(u', v')$, say $z = z(x, y)$, where $(x, y, z) = X$).

Since this lemma is known (cf. [8], p. 163, the last paragraph of Section 14), the proof of (ii bis) is now complete.

Remark. A corollary of (ii) is the following statement: If S is a surface of class C'', then it possesses C'-parametrizations $X = X(u^*; v^*)$ in terms of which the normal $N = N(u^*; v^*)$ is a function of class C'', too, provided that (45) holds on S. Since no mention is made now of normal parametrizations, that is, since it is not required that $[\partial N/\partial u^*, \partial N/\partial u^*] \neq 0$, it is natural to ask whether the proviso (45) can now be omitted. The answer proves to be in the negative. In fact, an example to this effect is supplied by every surface S which is of class C'' without being of class C''' (cf. the case $n = 2$ of the above Lemma) and which is a torse containing no flat points (that is, if $K \equiv 0 \neq H$ on S). Cf. Part V of [8].

Part III.

6. The purpose of the following theorems is to remove from the Codazzi equations belonging to the third fundamental form (Weingarten) unnecessary assumptions of differentiability which are implicit in the standard treatment of this problem. (cf. [1], pp. 232-234).

(iii) *If S is a surface of class C'' having no parabolic points and if $X = X(u, v)$ is a normal C'-parametrization of S on a simply connected domain D, then, in these parameters, the second and third fundamental forms*

(46) $\quad b_{ik}(u^1, u^2)\, du^i du^k, \qquad\qquad$ (47) $\quad c_{ik}(u^1, u^2)\, du^i du^k,$

given by (23), (24) *and* (15), *have the following properties:*

(†) *The form* (47) *is a C'-metric possessing the constant curvature $K_0 = 1$; the coefficients b_{ik} in* (46) *are of class C^0* (*continuous*) *and satisfy*

$$\det b_{ik} \neq 0; \tag{48}$$

finally, if the Γ^j_{ik} *are the Christoffel symbols of the second kind belonging to* (47), *then both relations*

$$\int_J b_{ik}du^k = \iint_E (\Gamma^j_{i1}b_{j2} - \Gamma^j_{i2}b_{j1})\,du^1du^2 \qquad (i = 1, 2) \tag{49}$$

are identities in J, *where* $E = E(J)$ *denotes the interior of any positively oriented, piecewise smooth Jordan curve* J *contained in* D.

In the classical statement of this theorem, it is (tacitly) supposed that S is of class C''''. If $X = X(u, v)$ is a C''''-parametrization of S, the matrices α, β, γ, defined by (16)-(21), are of class C''', C'', C'', respectively. In such a parametrization, the condition that (47) have the constant curvature $K_0 = 1$ is then expressible in terms of the second derivatives of the c_{ik}, while the "Codazzi" relations (49) are expressible in terms of the first derivatives of the b_{ik} (and c_{ik}). Thus (iii) improves on the classical theory by two degrees of differentiability—one resulting from the choice of normal parameters, and another from the fact that the condition on the curvature of (47) and the Codazzi relations (49) are used in an "integrated" form.

In a certain sense, conditions (†) of (iii) characterize the second and third fundamental forms in normal parametrizations of a surface of class C''. As does (iii), the following version of the "converse" of (iii) improves on the classical version by two degrees of differentiability.

(iii*) *Let the quadratic differential forms* (46), (47) *have the properties* (†) *of* (iii) *on a simply connected domain* D. *Then there exists a surface* S *of class* C'' *having on* D *a normal* C'*-parametrization* $X = X(u, v)$ *in which* (46) *and* (47) *become the second and third fundamental forms, respectively; that is,* (15) *is of class* C'' *and* (23), (24) *hold on* D. *The surface* S *is uniquely determined (up to movements of the Euclidean* X*-space).*

The proofs of (iii), (iii*) will only be sketched. They are modifications of the proofs in the classical case of higher differentiability (cf. [1], pp. 232-235). The necessary modification of those proofs is similar to the modification of the proof of Bonnet's theorem (characterizing the first and second fundamental forms), used in [6], pp. 761-762; the main point being that the standard theorem on total systems in the classical proofs is replaced by the theorem (II) of [6], pp. 760-761.

If $N = N(u, v)$, where $(u, v) = (u^1, u^2)$ is on a simply connected domain D, is a C''-parametrization of a portion of the sphere $|N| = 1$, then the derivation formulae of Gauss for the sphere are

(50) $$N_{ik} = \Gamma^j_{ik} N_j - c_{ik} N,$$

where c_{ik} is defined by (22) and Γ^j_{ik} are the Christoffel symbols belonging to (47). In view of (30), the Weingarten derivation formulae (29) for S can be written as

(51) $$X_k = - b_{ki} c^{ij} N_j, \quad \text{where} \quad k = 1, 2,$$

if $(c^{ij}) = \gamma^{-1}$, and if (45) and (48) hold.

If (50)-(51) is considered as a linear system for the unknowns X, N, N_1, N_2, then theorem (II) in [6,] pp. 760-761, shows that this system is "total" if and only if (47) has the constant curvature $K_0 = 1$ and the relations (49) hold as an identity in J. In fact, X, N, N_1, N_2 can be considered as scalars, since (50)-(51) consists of three identical systems, one for each component of the vectors. In this case (of four unknown functions and two independent variables), there are 16 $(= 4^2)$ integrability functions to be treated. When N is a scalar, (50) is identical with the Weingarten equation (13) if g_{ik} in the latter equation is replaced by c_{ik} and K_0 by 1. Hence, 9 of the 16 integrability conditions have been dealt with in the proof of (i), (i bis) and are satisfied if and only if (47) has the constant curvature $K_0 = 1$. Of the 7 remaining conditions, a direct calculation shows that 5 are trivial (of the type $0 = 0$) and 2 reduce to the relations (49).

Let S satisfy the conditions of (iii). Then the given parametrization $X = X(u, v)$ of S supplies an N and an X satisfying (50), (51). Every surface obtained from S by Euclidean movements also gives a solution X, N of (50), (51). By considering (50), (51) as scalar equations (say, as equations involving the first component), it is clear that from these solutions it is possible to construct solutions in which X, N, N_1, N_2 take arbitrary values at a given point of D, since (50)-(51) are linear and homogeneous. Hence, the necessity of the integrability conditions follows from theorem (II) of [6]. Since S has no parabolic points, (48) holds. Thus (46), (47) satisfy (†). This proves (iii).

In order to prove (iii*), note that if (46), (47) satisfy (†), then the "total" character of (50)-(51) leads to solutions X, N of class C', C'', respectively, uniquely determined by initial conditions. As in the classical case, a solution X, N satisfies (15), (19), (20) and (21), provided that the initial conditions of X, N, N_1, N_2 do. Finally, the Lemma (above) shows that there exists a surface S of class C'' for which X is a normal C'-parametrization. This proves (iii*).

A corollary of the proof of (iii*) is the fact that if conditions (†) are unaltered except that, in addition, it is supposed that (b_{ik}) is of class C',

then there result surfaces S of class C''' (instead of class C''), for which (46), (47) become the second and third fundamental forms in a normal C''-parametrization of S. The fact that no additional smoothness hypothesis is made on (47) seems curious at first glance. Actually, (i) shows that there is no loss of generality in assuming that (47) is analytic. A more inclusive consequence of the proof of (iii*) is as follows:

(iii$^*_n$) *If the forms* (46), (47) *are of class* C^n (*or* C^{n-1}), C^n *respectively, where* $n \geqq 1$, *and have the properties* (†) *of* (iii) *on a simply connected domain* D, *then there exists a surface* S *of class* C^{n+2} (*or* C^{n+1}) *having on* D *a* C^n- (*or* C^{n-1}-) *parametrization* $X = X(u, v)$ *in which* (15) *is of class* C^{n+1} *and for which* (20), (21) *hold.*

Part IV.

7. The point coordinates $X = (x, y, z)$ of S will now be replaced by its plane coordinates or its *supporting function*

$$w = X \cdot N \tag{52}$$

(which, according to (15), is defined even if S is of class C' only). In view of (iii), the result of the Appendix of [21], pp. 374-376, can be stated as follows:

(iv) *If* S *is a surface of class* C'' *without parabolic points and if* $X = X(\lambda, \mu)$ *is a spherical* C'-*parametrization of* S, *then the supporting function*

$$w = w(\lambda, \mu) = X(\lambda, \mu) \cdot N(\lambda, \mu) \tag{53}$$

is of class C'' (even though $X(\lambda, \mu)$ is of class C', and cannot be of class C'' unless S happens to be of class C''').

The proof of (iv) implies the following extension of (iv):

(iv$_n$) *The assertion of* (iv) *remains true if the respective classes* C'', C' *are replaced by* C^n, C^{n-1}, *where* $n \geqq 2$.

If S is of class C'' without parabolic points, then, by (iii), it posseses spherical C'-parametrizations $X = X(\lambda, \mu)$. After a suitable rotation of the X-space, it can be supposed that, in such a parametrization, (42)-(44) hold, where $\rho^2 < 1$. In this case, the notation

$$\lambda = \lambda^1, \qquad \mu = \lambda^2 \tag{54}$$

will be used and the letters a, b, c in (22), (23), (24) will be changed to g, h, f; that is, the three fundamental forms on S in terms of the (spherical) parameters (54) will be denoted by

$$(55) \quad |dX|^2 = g_{ik}d\lambda^i d\lambda^k; \qquad (56) \quad -dX \cdot dN = h_{ik}d\lambda^i d\lambda^k;$$

$$(57) \quad |dN|^2 = f_{ik}d\lambda^i d\lambda^k$$

(the representation $-X_i \cdot N_k = -X_k \cdot N_i$ of h_{ik} in (56) can, as in (20), be replaced by $X_{ik} \cdot N = h_{ik}$ only if not merely the surface S but also its C'-parametrization $X = X(\lambda, \mu)$ is of class C''; cf. the remark following (24) above).

Substitution of (42), (43), (44) into (57) shows that the coefficients f_{ik} of the third fundamental form are

$$(58) \qquad f_{11} = (1-\mu^2)/\nu^2, \qquad f_{12} = \lambda\mu/\nu^2, \qquad f_{22} = (1-\lambda^2)/\nu^2,$$

where $\nu^2 = 1 - \lambda^2 - \mu^2$. Since the function (53) is of class C'', it has (continuous) second covariant derivatives

$$(59) \qquad \nabla_{ik}w \equiv w_{ik} - \Gamma^j{}_{ik}w_j \qquad (i, k = 1, 2),$$

with reference to the Christoffel coefficients of the metric (57) defined by (58). A simple calculation shows that (59) becomes

$$(60) \qquad \nabla_{11}w = w_{11} - (1-\mu^2)\nu^{-2}\omega, \qquad \nabla_{12}w = w_{12} - \lambda\mu\nu^{-2}\omega,$$
$$\nabla_{22}w = w_{22} - (1-\lambda^2)\nu^{-2}\omega,$$

where

$$(61) \qquad \omega = \lambda w_1 + \mu w_2 \quad \text{and} \quad \nu^2 = 1 - \lambda^2 - \mu^2.$$

The subscripts of w in (59), (60), (61) (and later on) denote partial differentiations with respect to $\lambda = \lambda^1$ and $\mu = \lambda^2$.

In addition to (59), the following notations (cf. [1], p. 87) will be used: $\Delta_{22}w$ will denote the Monge-Ampère operator det $\{(f_{ik})^{-1}(\nabla_{ik}w)\}$ and $\Delta_2 w$ will be the Laplacian operator with respect to (57). In view of (58), this means that

$$(62) \qquad \Delta_{22}w = \nu^2 \det \nabla_{ik}w, \qquad (63) \qquad \Delta_2 w = \nu^2\Delta^2 w,$$

where

$$(64) \qquad \Delta^2 w = (1-\lambda^2)w_{11} - 2\lambda\mu w_{12} + (1-\mu^2)w_{22}.$$

Under the assumptions of (iv), the mean curvature H is continuous and the Gaussian curvature K is continuous and non-vanishing. In terms of the notations (62)-(64), the product and the sum of the principal radii of

curvature, that is, $1/K$ and $2H/K$, are given by Weingarten's formulae [17]

(65) $1/K = \Delta_{22}w + w\Delta_2 w + w^2$, (66) $2H/K = \Delta_2 w + 2w$

(cf. *loc. cit.*, p. 259), while the coefficients of the second form (56) follow from (42), (56) and (53),

$$-h_{ik} = w_{ik} + wf_{ik} \tag{66}$$

(*ibid.*, top of p. 259). The coefficients g_{ik} of the first fundamental form can be represented in terms of (58) and (52) by inserting (65) and (66) into

$$Kg_{ik} = 2Hw_{ik} + (1 + 2wH)f_{ik}. \tag{67}$$

In fact, (66) shows that (67) follows from (30) and (31), where α, β, γ are the matrices of (55), (56), (57), respectively.

Finally, the spherical parametrization $X = X(\lambda, \mu)$ of S can be obtained in terms of its supporting function $w = w(\lambda, \mu)$ and of (42), (58) as follows:

$$X = wN + f^{ik}w_i N_k, \quad \text{where} \quad (f^{ik}) = (f_{ik})^{-1} \tag{68}$$

(*ibid.*, p. 256 and p. 277).

Theorem (iv) and its consequence (68) have the following converse:

(iv*) *On a small domain of the circle $\lambda^2 + \mu^2 < 1$, let $w = w(\lambda, \mu)$ be a function of class C'' satisfying*

$$\Delta_{22}w + w\Delta_2 w + w^2 \neq 0 \tag{69}$$

(*cf.* (60)-(64)) *and let $N = N(\lambda, \mu)$ denote the unit vector defined by* (42)-(43). *Then there exists a unique surface S of class C'' having a spherical C'-parametrization $X = X(\lambda, \mu)$ with respect to which $\pm N(\lambda, \mu)$ is the normal* (15) *and $\pm w(\lambda, \mu)$ is the supporting function* (53) (*where $\pm$ is the signature of the expression* (69)).

It is clear from (65) that the condition (69) cannot be omitted in (iv*). The uniqueness of S is clear from the derivation of (68).

In order to prove (iv*), let $X = X(\lambda, \mu)$ be defined by (68) (and (58)). Clearly, (68) is a function of class C'. In view of (42), (58), (60) and (61), the vector (68) can be written as $X = (w - \omega)N + M$, where M is the "vector" $(w_1, w_2, 0)$. It is seen from (60) that $w_k - \omega_k = -\lambda^j w_{jk}$, where $k = 1, 2$; so that

$$X_k = (w - \omega)N_k - w_{jk}\lambda^j N + M_k.$$

Since the scalar product $M_k \cdot N$ is $w_{jk}\lambda^j$, it follows that X_1 and X_2 are orthogonal to N. Actually, the vector product $[X_1, X_2]$ is $\nu^{-1}N$ times the

expression in (69). (For the purposes at hand, it can be supposed that $(\lambda, \mu) = (0, 0)$ is a point of D, and it is then sufficient to verify the last assertion at that point.) Hence, when (69) holds, $X = X(\lambda, \mu)$ is a C'-parametrization of a surface S having $\pm N$ as its normal (15), where $\pm$ is the signature of the expression (69). It is clear from (68) that $\pm w$ is the supporting function of S. The Lemma (Section **5** above) implies that S is a surface of class C'', since both $X(\lambda, \mu)$ and $\pm M(\lambda, \mu)$ are of class C'.

It is clear from the proof that (iv*) can be extended as follows:

(iv$^*_n$) *The assertion* (iv*) *remains valid if the classes* C', C'' *are replaced by* C^n, C^{n+1}, *respectively, where* $n \geqq 1$.

Part V.

8. Theorem (iii*) concerns the "embedding" of a given pair of second and third fundamental forms. In what follows, there will be considered the embedding of a given (positive definite) third fundamental form and of either a given Gaussian curvature K $(\neq 0)$ or of a given mean curvature H; in other words, the embedding of a K $(\neq 0)$ or H given as a function of the normal N.

(v) *On a small simply connected* (u, v)*-domain* D, *let* (47) *be, for some* $n \geqq 1$, *a* C^n*-metric having the constant curvature* $K_0 = 1$. *Let* $\phi(u, v)$ *be a function of class* C^n *on* D. *Then there exist surfaces* S *of class* C^{n+1}, *having a normal* C^n*-parametrization* $X = X(u, v)$ *on* D, *with respect to which* (47) *is the third fundamental form* (24) *and the given* $\phi(u, v)$ *is any of the following functions:*

(v^1) *the mean curvature* $H(u, v)$;

(v^2) *the Gaussian curvature* $K(u, v)$, *provided that* $\phi \neq 0$;

(v^3) *the ratio* $2H/K$, *the sum of the principal radii of curvature.*

The auxilary condition $\phi \neq 0$ is necessary in (v^2), since it is assumed that (47) is positive definite. Since the metric (12), where $K_0 = 1$, and the form in (57) with coefficients (58) are equivalent by virtue of an analytic transformation $(u^*, v^*) \rightarrow (\lambda, \mu)$, it follows from (i) that it can be supposed that $(u, v) = (\lambda, \mu)$ and that the given coefficients c_{ik} are those given by (58). In fact, the C''-transformation $(u, v) \rightarrow (\lambda, \mu)$ defined by $(u, v) \rightarrow (u^*, v^*) \rightarrow (\lambda, \mu))$ will leave the assumptions of (†) on (47) and $\phi(u, v)$ unchanged. It can also be supposed that a given point (u, v) corresponds to $(\lambda, \mu) = (0, 0)$.

If $w = w(\lambda, \mu)$ is the supporting function of a surface S (the existence of which is to be proved), then, corresponding to (v^1), (v^2) or (v^3), the function w satisfies the respective partial differential equation

$$\Delta_2 w + 2w - 2(\Delta_{22} w + w\Delta_2 w + w^2)\phi(\lambda, \mu) = 0, \tag{70}$$

$$\Delta_{22} w + w\Delta_2 w + w^2 - 1/\phi(\lambda, \mu) = 0, \tag{71}$$

$$\Delta_2 w + 2w - \phi(\lambda, \mu) = 0; \tag{72}$$

cf. (65), (66). On the other hand, (iv^*), (iv^*_n) show that the embedding theorems (v^1), (v^2), (v^3) follow if it is verified that each of the partial differential equations (70), (71), (72) has, on a vicinity of $(\lambda, \mu) = (0, 0)$, solutions $w = w(\lambda, \mu)$ of class C^{n+1} satisfying (69).

Ad (72). It is clear from (63), (64) that the equation (72) is linear in the first and the second derivatives of w and is of elliptic type. Since ϕ is of class C^n, where $n \geqq 1$, it follows that (72) has solutions $w = w(\lambda, \mu)$ of class C^{n+1} (in a sufficiently small vicinity of $(\lambda, \mu) = (0, 0)$); cf. [15], pp. 91-92, and Section **9** below. Furthermore, solutions of (72) can be so chosen that (69) holds.

Remark. Since (72) is a linear, elliptic, partial differential equation, the above proof is valid if, instead of assuming that ϕ is of class C^n, it is only assumed that ϕ has $(n-1)$-st order partial derivatives which satisfy a uniform Hölder condition. When $n = 1$, this means that ϕ satisfies a uniform Hölder condition. On the other hand, it is indicated by considerations analogous to those concerning the Poisson and related equations (cf. [20]), that there exist continuous functions $\phi(\lambda, \mu)$ for which (72) has no solution on any (λ, μ)-domain. Hence, if $\phi(P)$ is an arbitrary continuous function of the position P on the sphere, it is unlikely that there exists a (closed) surface X of class C'' which has a normal image covering the sphere in a one-to-one manner and which satisfies $2H/K = \phi$, where the normal of X corresponds to the point P of the unit sphere. On the other hand the uniqueness of such a surface, if it exists, is known (Christoffel; cf. [12], p. 551).

Ad (70) *and* (71). The equations under consideration are of the Monge-Ampère type

$$A + Br + Cs + Dt + E(rt - s^2) = 0,$$

which is elliptic or hyperbolic according as

$$C^2 - 4BD + 4AE$$

is negative or positive.

In the case of (71), it is seen from (60)-(64) that, at $(\lambda, \mu) = (0, 0)$, the coefficients are $A = w^2 - 1/\phi$, $B = w$, $C = 0$, $D = w$ and $E = 1$. Hence, the expression in the last formula line becomes $-4/\phi$ at $(\lambda, \mu) = (0, 0)$. Thus (71) is elliptic or hyperbolic according as the assigned (non-vanishing) curvature ϕ is positive or negative. In the hyperbolic case, a known existence theorem (cf. [11], p. 849) shows that (71) has solutions of class C^{n+1} when ϕ is of class C^n, where $n \geqq 1$. In the elliptic case, the existence of such solutions can be deduced from an analogue of Lemma 2 of [7], p. 557; cf. Section **9** below. (In the case at hand, the condition (69) holds in view of (71)).

In the case of (70), the coefficients at $(\lambda, \mu) = (0, 0)$ are $A = 2\phi w^2 - 2w$, $B = 2\phi w - 1$, $C = 0$, $D = 2\phi w - 1$ and $E = 2\phi$. Hence, the expression in the last formula line becomes -4; so that (70) is of elliptic type near $(\lambda, \mu) = (0, 0)$. If ϕ is of class C^n, then (70) has solutions of class C^{n+1} satisfying (69). This is a consequence of the existence theorem in Section **9** below.

8 bis. Theorems (v^1), (v^2), (v^3) are special cases of an embedding theorem in which there are assigned as the third fundamental form a quadratic differential form (47) having the constant curvature $K_0 = 1$, and a given relation $F(2H/K, 1/K; u, v) = 0$ to be satisfied by $2H/K$ and $1/K$, the sum and the product of the principal radii of curvature. For the sake of simplicity, this general embedding theorem will be considered only when (u, v) is (λ, μ) and the given form (47) has the coefficients $c_{ik} = f_{ik}$ defined by (58). Let $F(U, V; \lambda, \mu)$ be a continuous function in a vicinity of a point $(U^0, V^0; 0, 0)$ and suppose that

$$(73) \qquad F(U^0, V^0; 0, 0) = 0, \quad V^0 \neq 0 \quad \text{and} \quad (U^0)^2 \geqq 4V^0.$$

The condition $V^0 \neq 0$ corresponds to the condition $1/K \neq 0$; the last part of (73) corresponds to the inequality $H^2 \geqq K$ and assures that (77) below can be satisfied by some numbers w^0, $w_1{}^0$, $w_{11}{}^0$, $w_{12}{}^0$, $w_{22}{}^0$. In addition, let F possess continuous partial derivatives F_U, F_V satisfying

$$(74) \qquad VF_V{}^2 + UF_UF_V + F_U{}^2 \neq 0.$$

In view of (65), (66) the embedding problem

$$(75) \qquad F(2H/K, 1/K; \lambda, \mu) = 0$$

depends on the partial differential equation

$$(76) \qquad F(\Delta_2 w + 2w, \Delta_{22} w + w\Delta_2 w + w^2; \lambda, \mu) = 0.$$

It is readily verified that, at the point $(U, V; \lambda, \mu) = (U^0, V^0; 0, 0)$, the equation (76) is of hyperbolic or elliptic type (that is, $4\partial F/\partial w_{11}\, \partial F/\partial w_{22} - (\partial F/\partial w_{12})^2$ is negative or positive) according as the expression (74) is negative or positive. By saying that (76) is of the hyperbolic or elliptic type at $(U^0, V^0; 0, 0)$, it is meant that if the left-hand side of (76) is considered to be a function of $(\lambda, \mu, w, w_1, w_2, w_{11}, w_{12}, w_{22})$, then (76) is of the specified type at a point $(0, 0, w^0, w_1{}^0, w_2{}^0, w_{11}{}^0, w_{12}{}^0, w_{22}{}^0)$ satisfying

$$(77)\qquad \begin{aligned} U^0 &= 2w^0 + w_{11}{}^0 + w_{22}{}^0, \\ V^0 &= w_{11}{}^0 w_{22}{}^0 - (w_{12}{}^0)^2 + (w^0)^2 + w^0(w_{11}{}^0 + w_{22}{}^0); \end{aligned}$$

cf. the remark following (73).

In order to assure the existence of a surface S, say of class C'', having a spherical C'-parametrization $X = X(\lambda, \mu)$, with respect to which $\pm (\lambda, \mu, \nu)$ is its normal vector and (75) is an identity in (λ, μ), where $H = H(\lambda, \mu)$, $K = K(\lambda, \mu) \neq 0$, additional requirements must be imposed on F. In the elliptic case, it is sufficient to require that $F = F(U, V; \lambda, \mu)$ be analytic with respect to (U, V) for fixed (λ, μ), that F and its partial derivatives of first and second order with respect to U, V satisfy a uniform Hölder condition with respect to their four arguments, and that the third derivatives of F with respect to U, V be continuous as functions of their four arguments); cf. Part VI below. In the hyperbolic cases, it is sufficient to require that $F(U, V; \lambda, \mu)$ be of class C'' with respect to its four variables together; cf. [11], pp. 847-848. In the particular hyperbolic case in which (76) is linear (in w_{11} w_{12}, w_{22}) or, more generally, of Monge-Ampère type, the assumption of the class C'' can be relaxed to the assumption that F is of class C'; cf. [11], pp. 848-849 and pp. 855-864.

Part VI.

9. The theorem referred to above, that on which the proof of (v) depends, is an existence theorem for solutions of an ellpitic partial differential equation on small domains. It is a generalization of the elliptic case of Lemma 2 in [7], p. 557, which is suggested by Picard's theorem [16] on the existence of solutions of linear boundary value problems on small domains and by Lichtenstein's theorem [15], p. 90, on non-linear elliptic partial differential equations involving a small parameter. The theorem in question, to be referred to as (§), is as follows:

(§) *Let $\Phi = \Phi(x, y, z, p, q, r, s, t)$ be a function on some eight-dimen-*

sional domain D which satisfies a uniform Hölder condition of order λ with respect to its eight variables, on every compact subset of D and, when (x, y) is fixed, let Φ be analytic with respect to the six variables (z, p, q, r, s, t). Let the first and second order partial derivatives of Φ with respect to z, p, q, r, s, t satisfy a uniform Hölder condition, of order λ, with respect to the eight variables (x, y, t, p, q, r, s, t), and let the third order partial derivatives of Φ with respect to z, p, q, r, s, t be continuous functions of (x, y, z, p, q, r, s, t). Finally, let

$$4\Phi_r\Phi_t - \Phi_s{}^2 > 0 \text{ on } S, \tag{78}$$

and let $P_0 = (x_0, y_0, z_0, p_0, q_0, r_0, s_0, t_0)$ be a point of D at which

$$\Phi(x, y, z, p, q, r, s, t) = 0. \tag{79}$$

Then, corresponding to a given neighborhood of P_0, there is a neighborhood of $(x, y) = (x_0, y_0)$ on which there exists a function $z = z(x, y)$ with the properties that z has second order partial derivatives satisfying a uniform Hölder condition of any order $\mu < \lambda$, the function $z(x, y)$ satisfies the partial differential equation (79), and (x, y, z, p, q, r, s, t) is within the given neighborhood of P_0.

The proof of (§), based on the method of successive approximations which depends on Korn's inequalities [14] in potential theory, will be essentially the same as that used in [18], pp. 64-68, and [15], pp. 90-98.

Proof of (§). Condition (78) implies that $\Phi_r \neq 0$ on P. Since (79) holds at P_0, it follows that (79) can be written in the form

$$r - \Psi(x, y, z, p, q, s, t) = 0 \tag{80}$$

in a neighborhood of P_0, where Ψ satisfies, on some seven-dimensional neighborhood of $(x_0, y_0, z_0, p_0, q_0, s_0, t_0)$, the smoothness conditions analogous to those satisfied by Φ on D. Consider the analytic partial differential equation

$$r - \Psi(x_0, y_0, z, p, q, s, t) = 0, \tag{81}$$

which is equivalent to $\Phi(x_0, y_0, z, p, r, s, t) = 0$. The equation (81) is satisfied at the point $(z, p, q, r, s, t) = (z_0, p_0, q_0, r_0, s_0, t_0)$. Hence, by the Cauchy-Kowalewski existence theorem, the assignment of analytic Cauchy data $z(x_0, y)$, $p(x_0, y)$ which, for $y = y_0$, reduce to $(z_0, p_0, q_0, s_0, t_0)$, where $q_0 = z_y(x_0, y_0)$, $s_0 = p_y(x_0, y_0)$, $t_0 = z_{yy}(x_0, y_0)$, determines a unique analytic solution $z = \zeta(x, y)$ of (81) in a neighborhood of (x_0, y_0). In particular, $\Phi(x_0, y_0, \zeta, \zeta_x, \cdots, \zeta_{yy}) \equiv 0$.

Put

$$z = \zeta(x, y) + u. \tag{82}$$

Then (79) becomes a partial differential equation for u and can be written in the form

$$(83)\qquad Au_{xx}+2Bu_{xy}+Cu_{yy}+Du_x+Eu_y+Fu=\Pi.$$

The coefficients A, B, C, D, E are the respective derivatives $\Phi_r, \Phi_s, \Phi_t, \Phi_p, \Phi_q, \Phi_z$ evaluated at $(x_0, y_0, \zeta, \zeta_x, \zeta_y, \zeta_{xx}, \zeta_{xy}, \zeta_{yy})$; so that A, B, C, D, E, F are analytic functions of (x, y). The function $\Pi=\Pi(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy})$ occurring in (83) is the difference of the left side of (83) and of $\Phi(x, y, \zeta+u, \cdots, \zeta_{yy}+u_{yy})$. The meaning of Π is clear from the fact that

$$\Phi(x, y, \zeta+u, \cdots, \zeta_{yy}+u_{yy})-\Phi(x_0, y_0, \zeta, \cdots, \zeta_{yy})$$
$$=Au_{xx}+Bu_{xy}+\cdots+Fu-\Pi.$$

Thus

$$(84)\qquad \Pi(x_0, y_0, 0, 0, \cdots, 0)=0,$$

and Π satisfies a uniform Hölder condition of order λ on a neighborhood of $(x, y, u_x, \cdots, u_{yy})=(x_0, y_0, 0, \cdots, 0)$. Also, Π is analytic, for fixed (x, y), in the independent variables $(u, u_x, \cdots, u_{yy})$, and has with respect to these variables first and second order partial derivatives which satisfy a uniform Hölder condition on a neighborhood of $(x_0, y_0, 0, \cdots, 0)$. In addition, Π has continuous third order derivatives with respect to the variables $(u, u_x, \cdots, u_{yy})$. The first order partial derivatives of Π satisfy

$$(85)\qquad \partial\Pi/\partial w=0 \text{ for } w=u, \cdots, u_{yy} \text{ at } (x, y, u, \cdots, u_{yy})=(x_0, y_0, 0, \cdots, 0).$$

Let $\delta>0$ be so small that $\zeta(x, y)$ is defined (and analytic) on the closure of

$$(86)\qquad \mathscr{R}_\delta:\quad (x-x_0)^2+(y-y_0)^2<\delta^2.$$

Let $m>0$ be chosen so that $\Pi(x, y, u, \cdots, u_{yy})$ has the properties enumerated, if (x, y) is in (86) and $(u, \cdots, u_{yy})$ is subject to the inequalities

$$(87)\qquad |u|\leqq m,\ |u_x|\leqq m, \cdots, |u_{yy}|\leqq m.$$

If $u(x, y)$ is any function on (86) satisfying a uniform Hölder condition of order λ, let $|u|_\lambda$ denote the least upper bound of the numbers M satisfying both inequalities

$$(88)\qquad |u(x, y)|\leqq M,\qquad |u(x+h, y+k)-u(x, y)|\leqq M(h^2+k^2)^{\frac{1}{2}\lambda}$$

for all (x, y), $(x+h, y+k)$ in (86). If $u(x, y)$ is of class $C''(\lambda)$ on (86) (that is, has second order partial derivatives on (86) which satisfy a uniform Hölder condition of order λ), put

$$\| u \|_\lambda = \max(| u |_\lambda, | u_x |_\lambda, \cdots, | u_{yy} |_\lambda). \tag{89}$$

Let N^0 denote an upper bound of the absolute value of Π and its first order partial derivatives with respect to $u, u_x, \cdots, u_{yy}$ on the product set of (86) and (87). Let $N \geqq N^0$ be an upper bound for the absolute value of the second and third order partial derivatives of Π on the domain specified. Finally, let N_λ denote an upper bound for the numbers M satisfying the inequality

$$| \Pi(x + h, y + k, u, \cdots, u_{yy}) - \Pi(x, y, u, \cdots, u_{yy}) | \leqq M(h^2 + k^2)^{\frac{1}{2}\lambda}, \tag{90}$$

and the corresponding inequalities which result if Π is replaced by any of its first and second partial derivatives with respect to $u, u_x, \cdots, u_{yy}$, for all (x, y), $(x + h, y + k)$ on (86), and $u, \cdots, u_{yy}$ satisfying (87).

For any $\epsilon > 0$, it follows from (84) and (85) that if $\delta > 0$ and $m > 0$ are sufficiently small, then N^0 can be chosen so as to satisfy

$$0 \leqq N^0 < \epsilon. \tag{91}$$

Also, if $0 < \mu < \lambda$ and if $\delta > 0$, $m > 0$ are sufficiently small, then N_μ can be chosen so as to satisfy

$$0 \leqq N_\mu < \delta \tag{92}$$

(in fact, $N_\lambda(2\delta)^{\lambda-\mu}$ is a possible choice for N_μ).

If $u = u(x, y)$ is a function of class $C''(\mu)$ on (86) satisfying (87), then the absolute value of $\Pi(x, y) \equiv \Pi(x, y, u(x, y) \cdots, u_{yy}(x, y))$ does not exceed N^0. Also, $| \Pi(x + h, y + k) - \Pi(x, y) |$ is not greater than the sum of

$$| \Pi(x + h, y + k, u(x + h, y + k), \cdots, u_{yy}(x + h, y + k)) \\ - \Pi(x, y, u(x + h, y + k), \cdots, u_{yy}(x + h, y + k)) |$$

and

$$| \Pi(x, y, u(x + h, y + k), \cdots, u_{yy}(x + h, y + k)) \\ - \Pi(x, y, u(x, y), \cdots, u_{yy}(x, y)) |.$$

This sum does not exceed

$$N_\mu(h^2 + k^2)^{\frac{1}{2}\mu} + N^0 | u(x + h, y + k) - u(x, y) | \\ + \cdots + N^0 | u_{yy}(x + h, y + k) - u_{yy}(x, y) |.$$

Hence

$$| \Pi(x, y) |_\mu \leqq N^0 + N_\mu + 6N^0 \| u \|_\mu. \tag{93}$$

If $u(x,y)$ and $u^*(x,y)$ form a pair of functions, of class $C''(\mu)$ on (86), satisfying

(94) $$\| u \|_\mu \leqq m \quad \text{and} \quad \| u^* \|_\mu \leqq m, \qquad (0 < m < 1),$$

and if $\Pi^*(x,y)$ denotes $\Pi(x,y,u^*(x,y),\cdots,u^*_{yy}(x,y))$, then there exist constants α, β, γ which are independent of δ and m (for small $\delta > 0$ and $m > 0$) and are such that

(95) $$|\Pi(x,y) - \Pi^*(x,y)|_\mu \leqq (\alpha N^0 + \beta N m + \gamma N_\lambda) \| u - u^* \|_\mu.$$

This can be proved by a device of Lichtenstein [15], pp. 93-94, as follows:

Let $\Phi(\tau) = \Phi(x,y,\zeta + \tau u,\cdots,\zeta_{yy} + \tau u_{yy})$ and let $\Phi^*(\tau)$ be defined analogously. Let $X(\tau) = \Phi(x_0,y_0,\zeta + \tau u,\cdots,\zeta_{yy} + \tau u_{yy})$ and let $X^*(\tau)$ be defined similarly. Then $-\Pi(x,y) = \Phi(1) - X'(0)$ and $-\Pi^*(x,y) = \Phi^*(1) - X^{*\prime}(0)$, where $' = d/d\tau$. Since $\Phi(0) - \Phi^*(0) = 0$,

(96) $$\Pi^* - \Pi = \int_0^1 \{(\Phi' - \Phi^{*\prime}) + X^{*\prime}(0) - X'(0)\} d\tau.$$

But the integrand in (96) can be written as

(97) $$(u - u^*)\{\Phi_z(x,y,\zeta + \tau u,\cdots) - A\} \\ + u^*\{\Phi_z(x,y,\zeta + \tau u,\cdots) - \Phi_z(x,y,\zeta + \tau u^*,\cdots)\} + \cdots,$$

where the last three dots indicate five more pairs of terms analogous to the pair displayed explicitly, and $A, B,\cdots,F$ are the coefficients of (83). Since the coefficient of $u - u^*$ is the partial derivative of $-\Pi$ with respect to z at the point $(x,y,\tau u,\cdots,\tau u_{yy})$, the integrand of (96) has the majorant $6N^0 \| u - u^* \|_\mu + 36 N \| u^* \|_\mu \| u - u^* \|_\mu$. Consequently, $|\Pi - \Pi^*|$ does not exceed a bound of the form occurring on the right-hand side of (95).

For any function $g = g(x,y)$, let $\Delta g = g(x+h, y+k) - g(x,y)$. Then, for fixed τ and $w = z, p,\cdots, s, t$,

(98) $$|\Delta \Phi_w(x,y,\zeta + \tau u,\cdots)| \leqq (N_\mu + 6N\tau \| u \|_\mu)(h^2 + k^2)^{\frac{1}{2}\mu};$$

cf. the derivation of (93). Also, for $h^2 + k^2 \neq 0$,

(99) $$(h^2 + k^2)^{-\frac{1}{2}\mu} |\Delta\{\Phi_w(x,y,\zeta + \tau u,\cdots) - \Phi_w(x,y,\zeta + \tau u^*,\cdots)\}|$$

does not exceed a bound of the form $(\alpha_1 N + \beta_1 N_\mu) \| u - u \|_\mu$, where α_1, β_1 are constants. In order to see this, let

$$\Phi^\tau(\sigma) = \Phi_w(x,y,\zeta + \tau u^* + \sigma\tau(u - u^*),\cdots).$$

Then $\Phi_w(x,y,\zeta + \tau u,\cdots) - \Phi_w(x,y,\zeta + \tau u^*,\cdots) = \Phi^\tau(1) - \Phi^\tau(0)$. But

$$d\Phi^\tau(\sigma)/d\sigma = \tau(u - u^*)\Phi_{wz} + (u_x - u^*_x)\Phi_{wp} + \cdots,$$

where the argument of $\Phi_{wz}, \Phi_{wp}, \cdots$ is $(x, y, \zeta + \tau u^* + \sigma\tau(u - u^*), \cdots)$. Since the third order partial derivatives of $\Phi(x, y, z, \cdots, t)$ with respect to $z, p, \cdots, t$ are bounded by N, it follows that $(h^2 + k^2)^{-\frac{1}{2}\mu}\Delta\{d\Phi^\tau(\sigma)/d\sigma\}$ has a bound of the type $(\alpha_2 N + \beta_2 N_\mu) \| u - u^* \|_\mu$ (cf. the derivation of (93)); hence the same holds for (99).

In view of (96), (97), (98) and the bound for (99), it is clear that there exist constants α, β, γ (independent of δ) such that (95) holds.

Theorem (§) can now be proved by the method of successive approximations. To this end, the following consequence of a classical result of Korn [14] will be needed: Let A, B, C, D, E, F be analytic functions of (x, y) on a set containing the circle (86), where $\delta > 0$ is sufficiently small, and let these functions satisfy

$$(100) \qquad B^2 - AC < 0.$$

Let $v(x, y)$ be defined on the closure of (86) and satisfy there a uniform Hölder condition of order μ. Then

$$(101) \qquad Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + F = v$$

has a unique solution $u = u(x, y)$ which is of class $C''(\mu)$ on the closure of $\mathcal{B}_\delta$ and vanishes on the boundary of $\mathcal{B}_\delta$. Furthermore, there exists a constant M depending on μ but independent of δ (for small δ) such that

$$(102) \qquad \| u \|_\mu \leqq M \mid v \mid_\mu, \qquad (M = M_\mu).$$

Let μ be any number satisfying $0 < \mu < \lambda$. Let A, B, C, D, E, F in (101) be identified with the coefficients of (83). With reference to the constant M in (102), let $\delta > 0$, $m > 0$ be fixed so small that

$$(103) \qquad M(N^0 + N_\mu + 6Nm) < m$$

and

$$(104) \qquad M(\alpha N^0 + \beta Nm + \gamma N_\mu) < \epsilon < 1.$$

This is clearly possible in view of the remarks concerning (91), (92) and of the independence of the constants α, β, γ in (95) of δ and m.

Define on the closure of $\mathcal{B}_\delta$ a sequence of functions $u^0, u^1, \cdots$ of class $C''(\mu)$ by induction, as follows: Let $u^0 \equiv 0$. Suppose that $u^0, u^1, \cdots, u^n$ have been defined, are of class $C''(\mu)$ on the closure of $\mathcal{B}_\delta$ and satisfy

$$(105) \qquad \| u^k \|_\mu \leqq m$$

for $k = 1, \cdots, n$. Then, by (93),

(106) $$|\,\Pi^n\,|_\mu \leqq N^0 + N_\mu + 6N^0 m,$$

where

(107) $$\Pi^n = \Pi^n(x, y) = \Pi(x, y, u^n, u^n{}_x, \cdots, u^n{}_{yy}).$$

Let $u = u^{n+1}$ be the unique solution, of class $C''(\mu)$, of

(108) $$Au_{xx} + \cdots + Fu = \Pi^n$$

which vanishes on the boundary of $\mathscr{B}_\delta$. Then u^{n+1} satisfies $\|\, u^{n+1}\,\|_\mu \leqq M\,|\,\Pi^n\,|$, by (102). Hence, (106) and (103) show that (105) holds for $k = n + 1$; so that the induction is complete.

The difference $u = u^{n+1} - u^n$ satisfies the partial differential equation

(109) $$Au_{xx} + \cdots + Fu = \Pi^n - \Pi^{n-1},$$

is of class $C''(\mu)$ on $\mathscr{B}_\delta$ and vanishes on the boundary of $\mathscr{B}_\delta$. Hence, by (102), $\|\, u^{n+1} - u^n\,\|_\mu \leqq M\,|\,\Pi^n - \Pi^{n-1}\,|_\mu$. According to (95),

$$|\,\Pi^n - \Pi^{n-1}\,|_\mu \leqq (\alpha N^0 + \beta N m + \gamma N_\mu)\,\|\, u^n - u^{n-1}\,\|_\mu.$$

Hence, by (104),

(110) $$\|\, u^{n+1} - u^n\,\|_\mu \leqq \epsilon\,\|\, u^n - u^{n-1}\,\|_\mu \qquad \text{for } n = 1, 2, \cdots.$$

It follows from $\epsilon < 1$ and from the definition of the norm $\|\cdots\|_\mu$ that the series $u^1 + (u^2 - u^1) + \cdots$ is absolutely and uniformly convergent on $\mathscr{B}_\delta$ to a function u of class $C''(\mu)$, and that this series can be differentiated term by term to obtain the first and second order partial derivatives of u.

Standard arguments show that $z = \zeta + u$ is a solution of (79) on $\mathscr{B}_\delta$. Since $u^1 - u^0 = u^1$ satisfies $\|\, u^1 - u^0\,\|_\mu \leqq m$, it follows from (110) that $\|\, u\,\|_\mu \leqq m/(1-\epsilon)$, that is, that $\|\, z - \zeta\,\|_\mu \leqq m/(1-\epsilon)$. In view of the fact that δ (hence m and ϵ) can be chosen arbitrarily small, and that $(x, y, \zeta, \zeta_x, \cdots, \zeta_{yy})$ reduces to $(x_0, y_0, z_0, p_0, \cdots, t_0)$ at $(x, y) = (x_0, y_0)$, the assertion (§) follows.

Part VII.

10. Let

(111) $$ds^2 = g_{ik}(u, v)\,du^i du^k, \qquad (u, v) = (u^1, u^2),$$

be a C'-metric on a vicinity D of $(u, v) = (0, 0)$ and let $\Gamma^j_{ik} = \Gamma^j_{ik}(u, v)$ be the corresponding Christoffel symbols. It is known ([3], p. 724) that given initial conditions need not determine a *unique* solution of

(112) $$u^{i\prime\prime} + \Gamma^i_{jk} u^{j\prime} u^{k\prime} = 0, \qquad i = 1, 2,$$

the differential equations of the geodesics of (111). The notion of a curvature of a C'-metric, described in Section **1** above, makes possible the formulation of the following theorem:

(I) *If* (111) *is a C'-metric on a (simply connected) vicinity D of $(u, v) = (0, 0)$ and possesses a curvature $K = K(u, v)$ which is a bounded function, then arbitrary initial conditions determine (locally) a unique solution of* (112).

This theorem is implied by the proof of theorem (I) in [3], p. 723. The assumptions of that theorem are quite different from those of (I) above; however, a perusal of the proof ([3], pp. 724-726) shows that it depends merely on the existence of a bounded curvature $K(u, v)$ in the sense of Section **1** above.

Under the assumptions of (I), there belongs to every ϕ a unique geodesic $J = J_\phi$, say

$$u = u(r, \phi), \quad v = v(r, \phi), \qquad (0 \leqq r \leqq \text{const.}, 0 \leqq \phi < 2\pi), \tag{113}$$

on which r is the arc-length,

$$u(0, \phi) = 0, \qquad v(0, \phi) = 0, \tag{114}$$

and ϕ is the angle between J_ϕ and a fixed direction at $(u, v) = (0, 0)$. The pair of functions (113) maps every sufficiently small circle ($0 \leqq r <$ const., $0 \leqq \phi < 2\pi$) in a continuous one-to-one manner on a vicinity of $(u, v) = (0, 0)$.

If the assumption of a *bounded* curvature for (111) is strengthened to the assumption of a *continuous* curvature, then the assertion of (I) can be strengthened as follows:

(II) *In addition to the assumptions of* (I), *let it be assumed that the curvature $K = K(u, v)$ of* (111) *is continuous. Let* (113) *be the unique geodesic satisfying* (114) *and making the angle ϕ with a fixed direction at $(u, v) = (0, 0)$, and let r be the arc-length on the geodesic* (113). *Then both functions* (113) *as well as their partial derivatives $u_r(r, \phi)$, $v_r(r, \phi)$ are of class C', and every sufficiently small circle* ($0 \leqq r \leqq$ const., $0 \leqq \phi < 2\pi$) *is mapped by* (113) *onto a vicinity of $(u, v) = (0, 0)$ in a one-to-one continuous manner and in such a way that, by virtue of this mapping,* (111) *becomes of the form*

$$ds^2 = dr^2 + g^2 d\phi^2, \tag{115}$$

where $g = g(r, \phi)$ is a continuous function possessing a continuous partial

derivative $g_r = g_r(r, \phi)$ *and satisfying*

$$(115\text{ bis}) \qquad g(0, \phi) \equiv 0, \qquad g_r(0, \phi) \equiv 1, \text{ and } g(r, \phi) > 0 \text{ if } r > 0.$$

This theorem is implied by the proof of theorem (III) of [9], p. 133, adapted from the proof of theorem 2 of [3], p. 724. It is easily seen that the proof of theorem (III) in [9], pp. 139-143, depends only on the existence of a *continuous* curvature, which proves (II) above.

It should be mentioned here that the wording of theorem (III) in [9] is erroneous; a corrected version is given by (II) above. The assumptions of theorem (III) in [9] imply only that (111) has a bounded curvature K, whereas its proof (cf. *loc. cit.*, the bottom of p. 141 and top of p. 142) requires somewhat more; for example, the continuity of K. In Section **12** below, there will be given an example showing that theorem (III) in [9] is false and that the assumption of (I) above (that is, the existence and the boundedness of K) do not imply the assertion of (II) above, claiming that (113) must be of class C'.

11. Under the assumptions of (II) above, it is possible to make an additional statement concerning the function g in (114), as follows:

LEMMA. *Under the assumptions of* (II), *the function* $g = g(r, \phi)$ *in* (114) *has a continuous second partial derivative* $g_{rr}(r, \phi)$ *with respect to* r *and satisfies the Jacobi equation*

$$(116) \qquad g_{rr} + Kg = 0, \qquad r > 0,$$

where $K = K(u, v)$ *is considered as a function of* (r, ϕ) *by virtue of* (113).

Essentially, this lemma was announced in [3], where a proof is indicated. In view of the difficulties involved in the details, a complete verification of the Lemma will be given here.

For a fixed (small) $r = c > 0$, (113) is a Jordan curve J^c of class C'. In fact, it is of class C'', although (113) need not be a C''-parametrization of J^r (cf. [4], or the example in Section **12** below). In order to show that J^c is of class C'', consider the inverse of the transformation (113),

$$(117) \qquad r = r(u, v), \qquad \phi = \phi(u, v).$$

The transformation (117) is of class C' and has a non-vanishing Jacobian in a punctured vicinity $(0 < u^2 + v^2 < \text{const.})$ of $(u, v) = (0, 0)$. The transformation rule of the contravariant form of the tensors occurring in (111)

and (115) implies, on this punctured vicinity, the identity

$$(118) \qquad g^{ik} r_i r_k = 1, \qquad (u^2 + v^2 > 0),$$

where $(g^{ik}) = (g_{ik})^{-1}$ and $r_1 = \partial r/\partial u$, $r_2 = \partial r/\partial v$. Locally, the partial differential equation (118) for $r = r(u, v)$ can be written in the form

$$(118\text{ bis}) \qquad r_u = F(r_v; u, v),$$

where F is analytic in r_v for fixed (u, v), and F and its partial derivatives with respect to r_v are of class C' in $(r_v; u, v)$.

Consider a point $(r, \phi) = (c, \phi^0)$ on the arc J^c. It can be supposed that the geodesic arc $\phi = \phi^0$ in (113) is the arc $u = 0$, for small $v \geqq 0$. In fact, the geodesic arc consisting of the geodesics $\phi = \phi^0$ and $\phi = \phi^0 + \pi$ in (113) has, for small $|v|$, a parametrization of the form $u = u(v)$, and the change of parameters $(u, v) \to (u - u(v), \pm v)$ is of class C'', has a non-vanishing Jacobian, leaves the assumptions of (II) invariant and transforms the arc consisting of $\phi = \phi^0$ and $\phi = \phi^0 + \pi$ into $u = 0$.

On the geodesic $u = 0$, the arc-length $r = r(0, v)$ is of class C'' as a function of v. The standard existence and uniqueness theorems for the partial differential equation of type (118 bis) show that $r = r(u, v)$ is of class C'' in a vicinity of the point $(u, v) = (0, v^0)$, corresponding to $(r, \phi) = (c, \phi^0)$. Since $r_u^2 + r_v^2 \neq 0$ at $(u, v) = (0, v^0) \neq (0, 0)$, it follows that the neighborhood of $(0, v^0)$ on the curve J^c: $r(u, v) = c$ is an arc of class C'', and so, since (c, ϕ^0) is an arbitrary point on the Jordan arc J^c, the latter is a curve of class C''.

Since (111) has a curvature K, the formula of Gauss-Bonnet,

$$(119) \qquad \int_J \kappa ds + \Sigma\, a_k + \iint_E K(g_{11}g_{22} - g_{12}^2)^{\frac{1}{2}}\, du dv = 2\pi,$$

is applicable to every region E bounded by a Jordan curve J which is piecewise of class C'' (cf. Section **1** above). Hence, an admissible domain $E = E(J)$ is

$$(120) \qquad E\colon\ (0 <)\, r_1 \leqq r \leqq r_2, \qquad \phi_1 \leqq \phi \leqq \phi_2$$

if r_2 is sufficiently small and $\phi_2 - \phi_1 < 2\pi$. In fact, the boundary J of (120) consists of two geodesic arcs, $\phi = \phi_1$ and $\phi = \phi_2$, where $r_1 \leqq r \leqq r_2$ (hence $r \neq 0$) and of arcs of two orthogonal trajectories J^c, where $c = r_1, r_2$. Clearly, $\Sigma a_k = 2\pi$ in the case (120) of (119). Hence the change of the integration variables u, v to r, ϕ in the case (120) of (119) gives

$$\int_J \kappa g ds = \iint_E K g dr d\phi, \tag{121}$$

since $ds = gd\phi$ and $g^2 = (g_{11}g_{22} - g_{12})^2(\partial(u,v)/\partial(r,\phi))^2$. On letting ϕ_1 and ϕ_2 tend to ϕ, and then r_1 and r_2 to r (> 0), it follows from the continuity of K and g that $\kappa = \kappa(r,\phi)$, the geodesic curvature of J^r at (r,ϕ), has with respect to r a partial derivative satisfying

$$(\kappa g)_r + Kg = 0. \tag{122}$$

Hence, in order to prove the Lemma, it remains only to show that the geodesic curvature κ satisfies

$$\kappa g = g_r. \tag{123}$$

Strictly speaking, κ in (119) is defined (on smooth sub-arcs of J) as $d\theta/ds$, where $-\theta$ is a (continuous) angle from a vector tangent to J to a vector in a field which is parallel in the sense of Levi-Civita along J; on the other hand, the geodesic curvature is $|\, d\theta/ds \,|$.

If (113) represents a C''-parametrization of J^c, so that g in (115) is of class C', then (123) follows from standard formulae for $d\theta/ds$ (cf., e. g., [1], p. 267). The point in the relation (123) is that (123) holds despite the fact that g need not be of class C'.

In order to prove (123) at a point $(r,\phi) = (c,\phi^0)$, where $c > 0$ is small, it can be supposed that $\phi^0 = 0$. Put

$$\psi = \psi(\phi) = \int_0^\phi g(c,t)\,dt \tag{124}$$

for ϕ near 0. Then ψ is of class C' (since g is continuous) and $\psi_\phi = g \neq 0$. Hence the function (124) has an inverse of class C',

$$\phi = \phi(\psi), \tag{125}$$

for ψ near 0. Consider the transformation $(u,v) \rightarrow (r,\psi)$,

$$u = u(r;\psi), \qquad v = r(r;\psi) \tag{126}$$

which results by substituting (125) into (113). Then (111) or (115) is transformed into

$$ds^2 = dr^2 + G d\psi^2, \quad \text{where} \quad G = g^2(r,\phi)/g^2(c,\phi) \tag{127}$$

is a continuous function of (r,ψ) near $(r,\psi) = (c, 0)$. The function G (as well as the other coefficients 1, 0 in (127)) can be calculated from the tensor transformation rule

(128) $$G = g_{ik}u^i{}_\psi u^k{}_\psi, \quad \text{where} \quad u^i = u^i(r;\psi),$$

and $u^1{}_\psi = \partial u/\partial\psi$, $u^2{}_\psi = \partial v/\partial\psi$ as well as the partial derivatives $u^i{}_r$, $u^i{}_\psi$, $u^i{}_{rr}$, $u^i{}_{r\psi} = u^i{}_{\psi r}$ exist and are continuous, by (II). In addition, $u^i{}_{\psi\psi}(c,\psi)$ exists and is continuous, since (124) shows that $r = c$ in (126) gives an arc-length parametrization of a portion of the curve J^c, which is of class C''. It follows that $G_r(r,\psi)$ and $G_\psi(c,\psi)$ exist and are continuous; moreover, these partial derivatives can be calculated from (128) by formal rules. Thus, although (127) is not a C'-metric, it has Christoffel symbols, say $\gamma^j{}_{ik}$, at the points (c,ψ), for small $|\psi|$, and these $\gamma^j{}_{ik}$ can be calculated in terms of the Christoffel symbols $\Gamma^j{}_{ik}$ of (111) by the standard transformation rule,

(129) $$\gamma^a{}_{jk}u^i{}_a = \Gamma^i{}_{\alpha\beta}u^\alpha{}_j u^\beta{}_k + u^i{}_{jk},$$

where $u^1 = u$, $u^2 = v$ are given by (126), the subscripts on u^i denote partial differentiation with respect to $v^2 = r$, $v^2 = \psi$, and the arguments of all functions correspond to $(r,\psi) = (c,\psi)$.

With reference to the parameters $u^1 = u$, $u^2 = v$, the differential equations defining parallel transport along J^c are

(130) $$W^{i\prime} + \Gamma^i{}_{jk}W^j u^{k\prime} = 0, \quad \text{where} \quad {}' = d/d\psi; \quad i = 1, 2.$$

In view of (129), this reduces to

(131) $$w^{i\prime} + \gamma^i{}_{jk}w^j v^{k\prime} = 0$$

by virtue of

(132) $$w^i = W^a v^i{}_a \quad \text{and/or} \quad W^i = w^a u^i{}_a.$$

In order to see this, note that $W^{i\prime} = w^{a\prime}u^i{}_a + wu^i{}_{a\beta}v^{\beta\prime}$ and that the second term of (130) is $\Gamma^i{}_{jk}w^a u^j{}_a u^k{}_\beta v^{\beta\prime}$; hence (131) follows from (129) if (130) is multiplied by $v^n{}_i$. Since $v^{1\prime} = 0$, $v^{2\prime} = 1$, the pair of equations (131) is $w^{i\prime} + \gamma^i{}_{j2}w^j = 0$. In view of $G(c,\psi) = 1$ (hence $G_\psi(c,\psi) = 0$), it follows that

$$\gamma^1{}_{12} = 0, \qquad \gamma^1{}_{22} = -\tfrac{1}{2}G_r(c,\psi) = -g_r(c,\phi)/g(c,\phi),$$

$$\gamma^2{}_{12} = \tfrac{1}{2}G_r = g_r(c,\phi)/g(c,\phi), \qquad \gamma^2{}_{22} = 0;$$

so that (131) becomes

(133) $$w^{1\prime} - (g_r/g)w^2 = 0, \qquad w^{2\prime} + (g_r/g)w^1 = 0.$$

In (123), $\kappa \doteq d\theta/ds$, where s is ψ and $\theta = \theta(s)$ is the angle (in the metric (111)) from any solution vector $(W^1(s), W^2(s)) \neq 0$ of (130) to the vector (u_ψ, v_ψ) tangent to J^c. Since length and angles are preserved under

transformations of class C' (on vectors and metrics), it follows from (132) that θ can be interpreted as the angle from any solution vector $(w^1(s), w^2(s)) \neq 0$ of (133) to the vector $(r_\psi, \psi_\psi) = (0, 1)$ tangent to J^c. Consider the solution of (133) satisfying the initial condition $w^1(0) = 0$, $w^2(0) = 1$. Then, if $\theta = \theta(s)$ is the angle from this solution vector to $(0, 1)$, it follows that $\sin\theta = w^1$, hence $d\theta/ds = w^{1\prime}$ at $s = 0$. Thus, the first equation in (133) and $w^2(0) = 1$ imply (123) at the point $(c, \phi^0) = (c, 0)$ of J^c. This proves the Lemma.

Since g in (115) need not be of class C' under the assumptions of (II), is is not possible to consider the differential equations of the geodesics in the (r, ϕ)-parameters. This defect in (II) can be remedied by an additional hypothesis on K, as follows:

(III) *In addtion to the assumptions of* (II), *let it be assumed that the curvature $K = K(u, v)$ of* (111) *is of class C'. Then the functions* (113) *are of class C'', and so the function g in* (115) *is of class C'.*

Assuming the Lemma above, an analogous theorem was proved in [5], p. 223, as follows: The assumption on K and the fact that (113) is of class C' imply that K in (116) is of class C' as a function of (r, ϕ). Hence, that solution $g = g(r, \phi)$ of the ordinary differential equation (116) which is deterimned by the initial conditions $g(0, \phi) = 0$, $g_r(0, \phi) = 1$ is of class C', and so (115) is a C'-metric for $r > 0$. It follows from the theorem (**) in [5], p. 222, that the functions occurring in (113) are of class C'' for small $u^2 + v^2 > 0$ (where $\partial(u, v)/\partial(r, \phi) \neq 0$). The difficulty at $(u, v) = (0, 0)$ can be eliminated by first considering u, v as functions, not of (r, ϕ), but of the Riemann coordinates $x = r\cos\phi$, $y = r\sin\phi$, and then verifying that (111) is transformed into a C'-metric in (dx, dy).

Clearly, (i) of Section **2** above can be deduced from (III).

(II_n)-(III_n) *Let* (111) *be a C^n-metric possessing a curvature $K = K(u, v)$ of class C^{n-1} or C^n, where $n \geqq 1$. Then the functions* (113) *are of class C^n, C^{n+1}, respectively, and g in* (115) *is correspondingly of class C^{n-1} or C^n.*

These theorems are proved in the same way as their particular cases, (II) and (III). (Actually, the cases $n \geqq 2$ are quite simple when compared to the cases $n = 1$).

12. There will now be given an example of a C'-metric (111) for which the partial derivatives $\partial g_{ik}/\partial u^j$ satisfy a uniform Lipschitz condition (which by [9], pp. 139-140, implies that there exists a bounded curvature

$K = K(u, v)$), but the corresponding functions (113) are not of class C'.

Such a metric is, for instance,

$$(134) \qquad ds^2 = h(v)(du^2 + dv^2), \text{ where } h(v) = 1 + \tfrac{1}{2}v^2 \operatorname{sgn} v,$$

if the domain D is the strip $|u| < \infty$, $|v| < 2^{\frac{1}{2}}$. If $v \neq 0$, it is readily calculated that (134) has the curvature

$$K = -\tfrac{1}{2}h^{-1}(\log h)_{vv} = -\tfrac{1}{2}(1 + \tfrac{1}{2}v^2 \operatorname{sgn} v)^{-3}(\operatorname{sgn} v - \tfrac{1}{2}v^2).$$

It follows that (134) has on D a curvature $K(u, v)$ which is bounded but not continuous ($K(u, +0) = \frac{1}{2}$, $K(u, -0) = \frac{1}{2}$). If u is used as a parameter, then (112) becomes

$$(135) \quad 2v''/(1 + v'^2) = (\log h)_v = |v|/(1 + \tfrac{1}{2}v^2 \operatorname{sgn} v), \text{ where } {}' = d/du;$$

cf. [1], p. 277. It is easily verified that the solution $v = v(u) = v(u; \phi)$ of (135) satisfying $v(0) = 0$ and $dv(0)/du = \tan \phi$ is given by

$$(136_1) \qquad v = 2^{\frac{1}{2}} \sinh(2^{-\frac{1}{2}}u/\cos \phi) \sin \phi \text{ for } u \geqq 0 \text{ when } 0 \leqq \phi < \tfrac{1}{2}\pi,$$

$$(136_2) \qquad v = 2^{\frac{1}{2}} \sin(2^{-\frac{1}{2}}u/\cos \phi) \sin \phi \text{ for } u \geqq 0 \text{ when } -\tfrac{1}{2}\pi < \phi < 0.$$

Hence $v(u; \phi)$ is continuous for $u \geqq 0$, $|\phi| < \frac{1}{2}\pi$. It is clear that $v(u; \phi)$ has a continuous partial derivative v_ϕ for $\phi \neq 0$ and that this derivative has the limits

$$(137_1) \qquad v_\phi(u; +0) = 2^{\frac{1}{2}} \sinh(2^{-\frac{1}{2}}u) \text{ for } u \geqq 0,$$

$$(137_2) \qquad v_\phi(u; -0) = 2^{\frac{1}{2}} \sin(2^{-\frac{1}{2}}u) \text{ for } u \geqq 0.$$

Thus, $v(u; \phi)$ is not of class C' (in fact, $v_\phi(u; 0)$ does not exist for small $u > 0$). It will be shown that, for this reason, the function $v = v(r, \phi)$ occurring in (113) cannot be class C'.

According to (134), the arc-length $r = r(u; \phi)$ on the geodesic (136_1) or (136_2) satisfies the relation $dr = h^{\frac{1}{2}}(v)(1 + (dv/du)^2)^{\frac{1}{2}}du$. Hence (136_1), (136_2) imply that

$$(138_1) \qquad r = (\cos \phi)^{-1} \int_0^u \{1 + \sinh^2(2^{-\frac{1}{2}}\tau/\cos \phi) \sin^2 \phi\} d\tau,$$

$$(138_2) \qquad r = (\cos \phi)^{-1} \int_0^u \{1 + \sin^2(2^{-\frac{1}{2}}\tau/\cos \phi) \sin^2 \phi\} d\tau,$$

respectively. It follows that $r(u; \phi)$ has a continuous partial derivative r_ϕ for $u \geqq 0$, $|\phi| < \frac{1}{2}\pi$. This is clear for $\phi > 0$ and $\phi < 0$ from (138_1) and

(138_2), respectively. If $\phi = 0$, it is sufficient to note that ϕ occurs in (138_1), (138_2) only in $\cos\phi$ and $\sin^2\phi$; so that $r_\phi(u;0)$ exists and is 0. Furthermore it is easily seen that $r_\phi(u;\phi)$ is continuous at $\phi = 0$ also. Hence, $r(u;\phi)$ is of class C'. Since $r_u \neq 0$, it follows that $r = r(u;\phi)$ can be solved for u and gives the function $u = u(r,\phi)$ in (113). Hence $u = u(r,\phi)$ is of class C'. The function $v = v(r,\phi)$ in (113) results by substituting $u = u(r,\phi)$ into $v = v(r,\phi)$. Thus, for $\phi \neq 0$, the function $v(r,\phi)$ has a continuous partial derivative,

$$v_\phi(r,\phi) = v_u(u;\phi)u_\phi(r,\phi) + v_\phi(u;\phi) \tag{139}$$

Since $v_u(u;\phi)$ is $\cosh(2^{-\frac{1}{2}}u/\cos\phi)\tan\phi$ or $\cos(2^{-\frac{1}{2}}u/\cos\phi)\tan\phi$ according as $\phi > 0$ or $\phi < 0$, it follows that $v_u(u;\phi) \to 0$ as $\phi \to 0$. But the continuity of $u_\phi(r,\phi)$ shows that $v_\phi(r,\phi)$ has the limits

$$v_\phi(r,+0) = v_\phi(u;+0) \text{ and } v_\phi(r,-0) = v_\phi(u;-0), \quad \text{where } u = u(r,\phi). \tag{140}$$

Hence it is seen from (137_1), (137_2) that the function $v(r,\phi)$ occurring in (113) is not of class C' (in fact, it has at $\phi = 0$ no partial derivative with respect to ϕ, for small $r > 0$.

Appendix.*

The object of this Appendix is to prove the following uniqueness theorem on closed convex surfaces:

($\square$) *Let $F(U,V;\lambda,\mu,\nu)$ be defined and of class C'' on the four-dimensional set*

$$U > 0,\ U^2 \geqq 4V > 0, \qquad \lambda^2 + \mu^2 + \nu^2 = 1$$

and satisfy

$$\operatorname{sgn} \partial F/\partial R_1 = \operatorname{sgn} \partial F/\partial R_2 \neq 0, \qquad U = R_1 + R_2,\ V = R_1R_2. \tag{1}$$

Then (up to translations) there is at most one closed surface S of class C''' with positive Gaussian curvature such that the principal radii of curvature R_1, R_2 at the point of S, where the inward unit normal vector is (λ,μ,ν), satisfies

$$F(R_1 + R_2, R_1R_2;\ \lambda,\mu,\nu) = 0. \tag{2}$$

---

* Added January 21, 1953.

When F is of the form $R_1 + R_2 - \phi(\lambda, \mu, \nu)$ or $R_1 R_2 - \phi(\lambda, \mu, \nu)$, then ($\square$) reduces to classical theorems (Christoffel, Minkowski), which are known to be valid under lighter smoothness assumptions on S, $\phi(\lambda, \mu, \nu)$ than what is assumed by ($\square$).

The assertion ($\square$) is due to Alexandroff [*1*] if F and S are restricted to be analytic. It has been proved by Pogoreloff [*5*] under the lighter hypothesis that F and S are of class C''' and C'''' (instead of C'' and C''', as in the above theorem), respectively. While the version of the theorem to be proved reduces by *one* the degree of differentiability assumed by Pogoreloff, it would be desirable to reduce the assumption of differentiability on S from C''' to C''; the class C'' being the natural one, the class of lowest differentiability in which R_1, R_2 are defined (by standard formulae) and are continuous. But the possibility of this reduction, $C''' \rightarrow C''$, will be left undecided.

By the assumption that F is of class C'' is meant that if, near any point of the sphere $\lambda^2 + \mu^2 + \nu^2 = 1$, one of the variables λ, μ, ν is expressed as a function of the other two (for example, $\nu = (1 - \lambda^2 - \mu^2)^{\frac{1}{2}} > 0$), then F is a function of class C'' of the four variables $(U, V; \lambda, \mu)$. Condition (*1*) means that the symbol $\neq$ in (74) can be replaced by $>$ and so the differential equation (76) is of elliptic type.

By using Theorem (†) in [*3*] (for a non-analytic elliptic equation), instead of the Lemma of Lewy [*4*], pp. 259-260 (for analytic elliptic equations), it is possible to prove ($\square$) by the procedure which was used by Lewy [*4*], pp. 261-262, to prove the analytic case of the theorem of Minkowski ($F = R_1 R_2 - \phi$) and which was adapted by Lewy from arguments applied by Cohn-Vossen [*2*], pp. 125-132, in an analogous problem.

Proof of ($\square$). Suppose that S and Σ are two surfaces (in an (x, y, z)-space) satisfying the conditions of ($\square$). It must be shown that Σ is a translation of S. According to (ii) and (ii bis) above, in a neighborhood of a point of the unit sphere, say of the point $(\lambda, \mu, \nu) = (0, 0, 1)$, the surfaces S and Σ, being of class C''', possess spherical C''-parametrizations $X(\lambda, \mu)$ and $\Xi(\lambda, \mu)$, respectively; so that (λ, μ, ν), where $\nu = (1 - \lambda^2 - \mu^2)^{\frac{1}{2}} > 0$, is the inward normal vector at the point $X(\lambda, \mu)$, $\Xi(\lambda, \mu)$ of S, Σ. Let $(h_{ik}(\lambda, \mu))$ and $(\eta_{ik}(\lambda, \mu))$ be the matrices of the second fundamental forms of S and Σ in terms of the parameters $(\lambda, \mu) = (\lambda^1, \lambda^2)$.

Consider, with Maxwell, those points (λ, μ), called by Cohn-Vossen "congruence points," at which $(h_{ik}) = (\eta_{ik})$. Then (*1*), (*2*) imply that either (λ, μ) is a congruence point or $\det(h_{ik} - \eta_{ik}) < 0$ at the point (λ, μ).

In order to see this, note that, by (33), the principal curvatures R_1, R_2 of S and Σ are the characteristic numbers of the matrix products $(f_{ik})^{-1}(h_{ik})$ and $(f_{ik})^{-1}(\eta_{ik})$, respectively, where (f_{ik}) is the common third fundamental matrix (58) of S and Σ. It is convenient to replace the matrix product $(f_{ik})^{-1}(h_{ik})$ by the positive definite symmetric matrix $(t_{ik}) = (f_{ik})^{-\frac{1}{2}}(h_{ik})(f_{ik})^{-\frac{1}{2}}$, having the same characteristic numbers, where $(f_{ik})^{\frac{1}{2}}$ is the (unique) positive definite square root of (f_{ik}). Similarly, $(f_{ik})^{-1}(\eta_{ik})$ can be replaced by $(\tau_{ik}) = (f_{ik})^{-\frac{1}{2}}(\eta_{ik})(f_{ik})^{-\frac{1}{2}}$. If R_1, R_2 ($\geqq R_1$) are the characteristic numbers of (t_{ik}) and ρ_1, ρ_2 ($\geqq \rho_1$) are those of (τ_{ik}), then (*1*) and (*2*) show that, unless $R_1 = \rho_1$ and $R_2 = \rho_2$, either $R_1 < \rho_1$ and $R_2 > \rho_2$ or $R_1 > \rho_1$ and $R_2 < \rho_2$. Hence, unless $(t_{ik}) = (\tau_{ik})$, it follows that $\det(t_{ik} - \tau_{ik}) < 0$ (as can be seen by considering the ellipses $t_{ik}d\lambda^i d\lambda^k = 1$ and $\tau_{ik}d\lambda^i d\lambda^k = 1$ in the $(d\lambda^1, d\lambda^2)$-plane, since these ellipses either coincide or intersect in four points; cf. [*2*], pp. 125-126, or [*4*], p. 262). This implies the disjunctive alternative: Either $(h_{ik}) = (\eta_{ik})$ or $\det(h_{ik} - \eta_{ik}) < 0$.

Accordingly, the (λ, μ)-net defined by

$$(3) \qquad (h_{ik} - \eta_{ik})\, d\lambda^i d\lambda^k = 0,$$

where $(\lambda^1, \lambda^2) = (\lambda, \mu)$, is real and has the congruence points as its singularities. It follows therefore by Cohn-Vossen's arguments in [*2*], pp. 126-127, that the uniqueness assertion of ($\square$) will be proved if it is shown that, unless S and Σ coincide after a translation, the congruence points are isolated and the index of any singular point of the net (*3*) is negative; cf. Lewy [*4*], pp. 261-262.

Let $w = X \cdot N$ and $\omega = \Xi \cdot N$ be the supporting functions (53) of S and Σ, respectively. Then $w(\lambda, \mu)$ and $\omega(\lambda, \mu)$ are of class C''', by (iv$_3$) above. Both w and ω satisfy the elliptic partial differential equation (*2*), where $R_1 + R_2 = 2H/K$ and $R_1R_2 = 1/K$ are given by (65), (66) and $\nu = (1 - \lambda^2 - \mu^2)^{\frac{1}{2}} > 0$. If $(\lambda, \mu, \nu) = (0, 0, 1)$ is a congruence point, it can be supposed that Σ has been translated so that $X(0,0) = \Xi(0,0)$. Then the difference $w - \omega$ and its partial derivatives $w_\lambda - \omega_\lambda$, $w_\mu - \omega_\mu$ are 0 at $(\lambda, \mu) = (0, 0)$. Theorem (†) in [*3*] implies that, unless S and Σ coincide, there exist an integer $n > 1$ and a negative constant c such that if $\chi = (\lambda^2 + \mu^2)^{\frac{1}{2}}$ and $\chi \to 0$,

$$w - \omega = o(\chi^n), \quad w_{ik} - \omega_{ik} = O(\chi^{n-1}) \text{ and } \det(w_{ik} - \omega_{ik}) \sim c\chi^{2n-2},$$

where the subscripts of w and ω denote partial differentiations with respect to $\lambda^1 = \lambda$, $\lambda^2 = \mu$.

If the last formula line is compared with (66), it is seen that

$$\det(h_{ik} - \eta_{ik}) \sim c\chi^{2n-2} \text{ as } \chi \to 0 \qquad (\chi^2 = \lambda^2 + \mu^2).$$

Since $c < 0$, the proof of the uniqueness theorem ($\square$) is now complete; cf. Lewy [4], p. 262.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] L. Bianchi, *Lezioni di geometria differenziale*, vol. I_1, 3rd ed. (1922).

[2] ———, *op. cit.*, II_2 (1924), pp. 480-488.

[3] P. Hartman, "On the local uniqueness of geodesics," *Ameriacn Journal of Mathematics*, vol. 72 (1950), pp. 723-730.

[4] ———, "On geodesic coordinates," *ibid.*, vol. 73 (1951), pp. 949-954.

[5] ———, "On unsmooth two-dimensional Riemannian metrics," *ibid.*, vol. 74 (1952), pp. 215-226.

[6] ——— and A. Wintner, "On the fundamental equations of differential geometry," *ibid.*, vol. 72 (1950), pp. 757-774.

[7] ——— and A. Wintner, "On the embedding problem in differential geometry," *ibid.*, vol. 72 (1950), pp. 553-564; cf. *ibid.*, vol. 74 (1952), p. 264.

[8] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[9] ——— and A. Wintner, "On the problems of geodesics in the small," *ibid.*, vol 73 (1951), pp. 132-148.

[10] ——— and A. Wintner, "Gaussian curvature and local embedding," *ibid.*, vol. 73 (1951), pp. 876-884.

[11] ——— and A. Wintner, "On hyperbolic differential equations, *ibid.*, vol. 74 (1952), pp. 834-864.

[12] A. Hurwitz, *Mathematische Werke*, vol. 1 (1932), pp. 548-554.

[13] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *American Journal of Mathematics*, vol. 60 (1938), pp. 129-138.

[14] A. Korn, "Zwei Anwendungen der Methode der sukzessiven Annäherungen," *Schwarz Festschrift* (1914), pp. 215-229.

[15] L. Lichtenstein, *Vorlesungen über einige Klassen nicht-linearer Integralgleichungen und Integro-Differentialgleichungen*, Berlin (1931).

[16] E. Picard, "Mémoire sur la théorie des équations aux dérivées partielles et la méthode des approximations successives," *Journal de Mathématiques*, ser. 4, vol. 6 (1890), pp. 145-210.

[17] J. Weingarten, "Ueber die Theorie der auf einander abwickelbaren Oberflächen," *Festschrift der Technischen Hochschule zu Berlin*, 1884.

[18] H. Weyl, "Ueber die Bestimmung einer geschlossen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[19] A. Wintner, "On Riemann metrics of constant curvature," *American Journal of Mathematics*, vol. 73 (1951), pp. 569-575.

[20] ———, "On the Hölder restrictions in the theory of partial differential equations," *ibid.*, vol. 72 (1950), pp. 731-738.

[21] ———, "On parallel surfaces," *ibid.*, vol. 74 (1952), pp. 365-376.

APPENDIX.

[*1*] A. D. Alexandroff, "Sur les théorèmes d'unicité pour les surfaces fermées," *Comptes Rendus (Doklady) URSS*, vol. 22 (1939), pp. 99-102.

[*2*] St. Cohn-Vossen, "Zwei Sätze über die Starrheit der Eiflächen," *Göttinger Nachrichten*, 1927, pp. 125-134.

[*3*] P. Hartman and A. Wintner, "On elliptic partial differential equations," *American Journal of Mathematics*, vol. 75 (1953), (to appear).

[*4*] H. Lewy, "On differential geometry in the large, I," *Transactions of the American Mathematical Society*, vol. 43 (1938), pp. 258-270.

[*5*] A. V. Pogoreloff, "Extension of a general uniqueness theorem of A. D. Alexandroff to the case of non-analytic surfaces," *Doklady Akademii Nauk SSSR (N. S.)*, vol. 62 (1948), pp. 297-299 (in Russian); cf. *Zentralblatt für Mathematik*, vol. 30 (1949), p. 414, and/or *Mathematical Reviews*, vol. 10 (1949), p. 325.

GENERALIZED ASYMPTOTIC DENSITY.*

By R. Creighton Buck.

1. Introduction. In this paper, we shall discuss the properties of an asymptotic density in the context of a general σ-finite measure space; in particular, we shall prove what may be called the additivity theorem. This was originally formulated and proved only in the very special case of asymptotic density of sets of integers [2]. This introduces a certain amount of unification into the theory of asymptotic density. In Section **3**, a number of applications will be made of the additivity theorem.

Let X be a set on which a density function is to be defined. Topological properties of X will play no role; however, the notion of boundedness is important. We suppose that a countable sequence of sets $K(1) \subset K(2) \subset \cdots$ has been chosen so that $\bigcup K(n) = X$. A set $S \subset X$ is called bounded if $S \subset K(n)$ for some n. Let μ_n be a sequence of measures on X, having in common a field $\mathfrak{M}$ of measurable sets which contains at least the sets $K(n)$ and X. We require two conditions: (i) $\mu_n(X) = 1$ for $n = 1, 2, \cdots$; (ii) $\mu_n(K(j)) \to 0$, as $n \to \infty$, for each $j = 1, 2, \cdots$. In all that follows, we shall be discussing sets in the class $\mathfrak{M}$. For certain of these, we define a density function $D(S)$ as $\lim \mu_n(S)$, when this limit exists. For any set S in $\mathfrak{M}$, upper and lower densities $\bar{D}(S)$ and $\underline{D}(S)$ are defined as the upper and lower limits of the sequence $\mu_n(S)$. Thus, S has density only when $\bar{D}(S)$ and $\underline{D}(S)$ agree.

Examples of densities are easily given. If ν is a measure on X for which $\nu(X) = +\infty$ while $\nu(K(j)) < \infty$ for all j, then ν defines a density on X with measures μ_n defined by $\mu_n(S) = \nu(S \cap K(n))/\nu(K(n))$. In particular, the customary asymptotic density defined for sets of integers is obtained by the choice of X as $\{1, 2, \cdots\}$ $K(n)$ as $\{1, 2, \cdots, n\}$ and ν as the point measure with mass 1 at each point.

The density function D with measures μ_n has many properties similar to that of a measure itself. These are summarized in the following statement:

1) $D(X) = 1$, and if S is bounded, $D(S) = 0$.

* Received November 10, 1952.

2) If S has density, so does $X - S$ and $D(X - S) = 1 - D(S)$.

3) If A and B have density, and $A \subset B$, then $D(A) \leqq D(B)$.

4) If A and B have density, and are disjoint, then $A \cup B$ has density, and $D(A \cup B) = D(A) + D(B)$.

However, in contrast, in 4) if A and B are not disjoint, then neither $A \cup B$ nor $A \cap B$ need have density. Moreover, D is not countably additive; for example, setting $J_n = K(n) - K(n-1)$ we have $D(J_n) = 0$ for all n while $D(\bigcup J_n) = D(X) = 1$. It is possible to overcome the latter defect by introducing a modified set inclusion and set union and thus obtain a true additivity theorem. The device is essentially that of working modulo bounded sets. We shall write $A \dot{\subset} B$ when $A - B$ is bounded; there is then a value of j for which $A - K(j) \subset B - K(j)$. If $A \dot{\subset} B$, then since bounded sets are of zero density, $\underline{D}(A) \leqq \underline{D}(B)$ and $\bar{D}(A) \leqq \bar{D}(B)$; thus, D is still monotone.

At this point we impose a further condition on the measures μ_n which is needed in the proof of the next theorem:

(iii) the measure μ_n has bounded support.

This requires that there exist a sequence $a(n)$ such that $\mu_n(S) = 0$ for any set S disjoint from $K(a(n))$. We observe that the density obtained from a single measure ν as described above obeys (iii); the support of μ_n is contained in $K(n)$ itself.

THEOREM 1.1. *Let $A_1 \dot{\subset} A_2 \dot{\subset} A_3 \dot{\subset} \cdots$ and let* $\lim \bar{D}(A_n) = \Delta$ *and* $\lim \underline{D}(A_n) = \delta$. *Then, there exists a set A with $\bar{D}(A) = \Delta$ and $\underline{D}(A) = \delta$, such that $A_n \dot{\subset} A$ for all n.*

COROLLARY 1. *If the sets A_n have density, and $A_1 \dot{\subset} A_2 \dot{\subset} \cdots$ then there is a set A, unique up to sets of zero density, such that $A_n \dot{\subset} A$ for all n, and $D(A) = \lim D(A_n)$.*

COROLLARY 2. *If $C_1, C_2, \cdots$ are disjoint sets having density, there is a set C, unique up to sets of zero density, such that $C \dot{\supset} \bigcup_1^n C_k$ for all n, and with $D(C) = \sum_1^\infty D(C_k)$.*

Proof of the theorem. Assume for the present that the sets A_n are monotone in the usual sense. Given $\epsilon > 0$, choose b_n so that $\mu_{b_n}(A_n) \leqq \underline{D}(A_n) + \epsilon$ and for all $k \geqq b_n$, $\mu_k(A_n) \leqq \bar{D}(A_n) + \epsilon$. With the sequence $a(n)$ having the meaning given above, let $I_1 = K(a(b_1))$, and $I_n = K(a(b_n)) - K(a(b_{n-1}))$.

The set A is defined by $A = \bigcup_1^\infty A_n \cap I_n$. Since $A_n \supset A_j$ for all $n \geqq j$, $A \cap \bigcup_{n \geqq j} I_n = \bigcup_{n \geqq j} A_n \cap I_n \supset \bigcap_{n \geqq j} A_j \cap I_n = A_j \cap \bigcup_{n \geqq j} I_n$ and $A \dot{\supset} A_j$. From the monotone character of upper and lower density, and the definition of Δ and δ, $\bar{D}(A) \geqq \Delta$ and $\underline{D}(A) \geqq \delta$. If $n \leqq j$, then $A_n \subset A_j$ so that

$$A \cap \bigcup_{n \leqq j} I_n = \bigcup_{n \leqq j} A_n \cap I_n \subset \bigcup A_j \cap I_n = A_j \cap \bigcup_{n \leqq j} I_n,$$

or $A \cap K(a(b_j)) \subset A_j \cap K(a(b_j))$. Since the support of μ_k lies in $K(a(k))$, $\mu_k(S) = \mu_k(S \cap K(a(b_j)))$ whenever $k \geqq b_j$. In particular, $\mu_k(A) \leqq \mu_k(A_j)$ when $k \geqq b_j$. Using the definition of b_j, we have $\mu_k(A) \leqq \bar{D}(A_j) + \epsilon$ for all $k \geqq b_j$, and $\mu_{b_j}(A) \leqq \underline{D}(A_j) + \epsilon$. This immediately yields $\bar{D}(A) \leqq \Delta + \epsilon$, and $\underline{D}(A) \leqq \delta + \epsilon$. Letting ϵ decrease, we see that $\bar{D}(A)$ and $\underline{D}(A)$ must equal Δ and δ respectively. Returning to the original hypothesis that $A_1 \dot{\subset} A_2 \dot{\subset} A_3 \dot{\subset} \cdots$, there exist bounded sets B_n such that $A_n - B_n \subset A_{n+1} - B_n$ for each n. Set $A_n^* = A_n \cup (B_1 \cup B_2 \cup \cdots \cup B_{n-1})$. Applying the previous argument to this sequence gives the desired result.

The only unproved statement in Corollary 1 is the uniqueness of A. If A^* is another set for which both $D(A^*) = \lim D(A_n)$ and $A_n \dot{\subset} A^*$ for all n, then so is the set $A \cap A^*$, and since $\mu_k(A) = \mu_k(A - A^*) + \mu_k(A \cap A^*)$, $D(A - A^*) = 0$. Likewise $D(A^* - A) = 0$, and A and A^* differ only by a set of zero density.

In connection with this theorem, there is a simpler result concerned with a single measure on X which, especially in its applications, has a certain similarity.

THEOREM 1.2. *Let ν be a measure on X, finite on bounded sets. Let $A_1 \subset A_2 \subset \cdots$ be a sequence of sets of finite measure. Then given $\delta > 0$, there is a set A with $\nu(A) < \delta$ and $A_n \dot{\subset} A$ for all n.*

Let $R_n = X - K(n)$. If B is of finite measure, then $\nu(B \cap R_n) \to 0$ as n increases. Let k_n be a value of k for which $\nu(A_n \cap R_k) < \delta/2^n$, and set $I_1 = K(k_1)$, $I_n = K(k_n) - K(k_{n-1})$. As before let $A = \bigcup A_n \cap I_n$; we again have $A \dot{\supset} A_j$ for all j. Computing $\nu(A)$, we have

$$\nu(A) \leqq \Sigma \nu(A_n \cap I_n) \leqq \Sigma \nu(A_n \cap R_{k_n}) \leqq \Sigma \delta/2^n = \delta.$$

The set A cannot be taken of zero measure unless each of the sets A_n is essentially bounded, in the sense that there is a bounded set B with $\nu(A - B) = 0$.

2. Applications. We shall be chiefly concerned with real valued functions defined on X, measurable, and bounded on bounded sets. In Section **4**, we shall discuss the extension to complex-valued functions. Although X has no topology, we can use the sets $R_n = X - K(n)$ to define one 'at infinity.' We say that $\lim f(x) = L$, as $x \to \infty$ in X, if, for any $\epsilon > 0$ there is an n such that $|f(x) - L| < \epsilon$ for all $x \varepsilon R_n$. Associated with the measures μ_n, a method of summability can be defined as follows: (μ)-$\lim f(x) = L$ as $x \to \infty$ if and only if $\lim \int f d\mu_n = L$ as $n \to \infty$.

THEOREM 2.1. *(μ)-summability is regular. If $\lim f(x) = L$ then (μ)-$\lim f(x) = L$ as $x \to \infty$, for L finite, or $+\infty$ or $-\infty$.*

If L is finite, we may take $L = 0$. Let $|f(x)| < \epsilon$ on R_j. Then,

$$\left| \int f \, d\mu_n \right| \leqq \int_{K_j} |f| \, d\mu_n + \int_{R_j} |f| \, d\mu_n \leqq O(1)\mu_n(K_j) + \epsilon\mu_n(R_j).$$

Hence, $\limsup |\int f \, d\mu_n| \leqq 0 + \epsilon$, and since ϵ is arbitrary, (μ)-$\lim f(x) = 0$ as $x \to \infty$. The remaining cases are similar.

The density of a set $S \subset X$ may be alternatively defined by means of this generalized limit operation. If S is any set having density, it is evident that $D(S) = (\mu)$-$\lim s(x)$ where $s(x)$ is the characteristic function of S. It is to be expected therefore that there should be close connections between these two. We shall say that $f(x)$ converges to L 'in density' if there is a set $A \subset X$ of zero density such that $\lim f(x) = L$, as $x \to \infty$ in $X - A$. The next few theorems will show the connection between this and summability of $f(x)$.

THEOREM 2.2. *If $f(x)$ is bounded on X, and converges to a finite limit L in density, then (μ)-$\lim f(x) = L$.*

Take $L = 0$ and suppose $|f(x)| < \epsilon$ on $R_j - A$, $|f(x)| < M$ on X. Then,

$$\left| \int f \, d\mu_n \right| \leqq \int_{K_j \cup A} |f| \, d\mu_n + \int_{R_j - A} |f| \, d\mu_n \leqq M\mu_n(K_j \cup A) + \epsilon\mu_n(R_j - A).$$

Since $D(A) = D(K_j) = 0$, we again have $\limsup |\int f \, d\mu_n| \leqq \epsilon$, so that (μ)-$\lim f(x) = 0$.

When L is infinite, the theorem takes a slightly different form.

THEOREM 2.3. *If $f(x)$ is bounded from below on X, and converges to $+\infty$ in density, then (μ)-$\lim f(x) = +\infty$.*

Suppose that $f(x) > C$ for $x \varepsilon R_j - A$, and $f(x) > -M$ on X. Then, as above $\int f\, d\mu_n > (-M)\mu_n(K_j \cup A) + C\mu_n(R_j - A)$ and $\liminf \int f\, d\mu_n \geqq C$. Since C may be taken arbitrarily large, $(\mu)\text{-}\lim f(x) = +\infty$.

The value of $(\mu)\text{-}\lim f(x)$ when it exists must lie between $\liminf f(x)$ and $\limsup f(x)$. Since it is in the nature of an average of the values of $f(x)$ when x is 'large,' it might be conjectured that if $(\mu)\text{-}\lim f(x)$ exists and is either of the end points of this interval, the values of $f(x)$ must themselves cluster around this value. The following theorem confirms this; the proof makes use of the additivity theorem.

THEOREM 2.4. *Let* $L = \liminf f(x)$ *be finite, and suppose that* $(\mu)\text{-}\lim f(x) = L$. *Then,* $f(x)$ *converges to* L *in density.*

We may take $L = 0$. For a given $\delta > 0$, let $A = \{x \mid f(x) > \delta\}$. Given $\epsilon > 0$, choose j so that $f(x) > -\epsilon$ on R_j. Since f is bounded on bounded sets, $|f(x)| < M$ for all x in $K(j)$. Then,

$$\int f\, d\mu_n \geqq -M\mu_n(K(j)) + \delta\mu_n(R_j \cap A) - \epsilon\mu_n(R_j),$$

as seen by splitting the integration range into $K(j)$, $R_j \cap A$, $R_j - A$. As n increases, we obtain $0 \geqq 0 + \delta\bar{D}(A) - \epsilon$ and since ϵ may be arbitrarily small, $\bar{D}(A) = D(A) = 0$. If we set $\delta = 1, 1/2, 1/3, \cdots, 1/n, \cdots$ the corresponding sets A_n form an increasing sequence of sets of zero density. Applying Theorem 1.1, let B be the modified union of the sets $\{A_n\}$, obeying the conditions: $D(B) = 0$ and $B \dot{\supset} A_n$ for all n. We show that $f(x)$ converges to 0 off B. Given ϵ, choose $n > 1/\epsilon$; since $B \dot{\supset} A_n$ there is an i with $B - K(i) \supset A_n - K(i)$. If $x \varepsilon R_i - B$ then $x \varepsilon A_n$ and $f(x) \leqq 1/n < \epsilon$. On R_j, $f(x) > -\epsilon$. If $k = \max{:}\, i, j{:}$, then $|f(x)| < \epsilon$ on $R_k - B$.

COROLLARY. *In this theorem, the hypothesis* $\liminf f(x) = L$ *can be relaxed to:* (i) $\inf_X f(x) > -\infty$, (ii) *for every* $\epsilon > 0$ *the set of* x *for which* $f(x) > L - \epsilon$ *has unit density.*

Again, we take $L = 0$. Let $A_\epsilon = \{x \mid f(x) > -\epsilon\}$; as ϵ decreases, these sets decrease, and by the additivity theorem, their modified intersection A is a set of unit density such that as $x \to \infty$ in A, $\liminf f(x) \geqq 0$. To prove the corollary, we apply the theorem with A replacing the whole space X. Let $f(x) > -M$ on X. Then,

$$\int_A f\, d\mu_n = \int_X f\, d\mu_n - \int_{X-A} f\, d\mu_n \leqq \int_X f\, d\mu_n + M\mu_n(X - A)$$

so that $\limsup \int_A f\,d\mu_n \leqq 0$. Since $\liminf_A f(x) \geqq 0$ we have both $\liminf_A f(x) = 0$ and $\lim \int_A f\,d\mu_n = 0$. Applying the theorem, there is a set $B \subset A$ of unit density in A and therefore of unit density in X, such that $f(x)$ converges to 0 as $x \to \infty$ in B.

In the next section, we shall give an example to show that condition (ii) by itself is not enough to imply the conclusion.

An analogous application of Theorem 1.2 can be made.

THEOREM 2.5. *Let ν be a measure on X, finite on bounded sets, and let $\int |f|\,d\nu < \infty$. Then, for any $\delta > 0$, there is a set A, $\nu(A) < \delta$, such that $f(x)$ converges to 0 as $x \to \infty$ in $X - A$.*

As before, let $A_n = \{x \mid |f(x)| > 1/n\}$. Then,

$$\nu(A_n) = n \int_{A_n} (1/n)\,d\nu \leqq n \int_{A_n} |f|\,d\nu \leqq n \int_X |f|\,d\nu < \infty.$$

The sets $\{A_n\}$ form an increasing sequence of sets of finite measure. By Theorem 1.2, there is a set A, $\nu(A) < \delta$, such that $A \dot{\supset} A_n$ for all n. From here on, the proof proceeds much as in the previous theorem.

3. Special applications.

In this section, we will be concerned with a number of illustrations which arise from results of the previous section by special choices of the measures μ_n and ν involved. We take X as $\{1, 2, 3, \cdots\}$. Let ν be the measure on X which assigns to the point n the mass $1/n$. For $S \subset X$, $\nu(S)$ is finite if and only if $\sum 1/n < \infty$, where $n \in S$. Thus, sets of finite ν measure are the same as sets of "zero logarithmic density." With these substitutions, Theorem 2.5 becomes:

THEOREM 3.1. *If $\sum |c_n| < \infty$, then nc_n converges to 0 except for a subsequence of zero logarithmic density.* (See also [3])

This may be used in the theory of Fourier series.

COROLLARY 1. *Let $\{\phi_n\}$ be orthonormal on $[0, 1]$. Let $\lambda_n \downarrow 0$ and let $\{c_n\}$ be a sequence satisfying $\sum \lambda_n |c_n|^2 < \infty$. Set $F_n(x) = \sum_0^n c_j\phi_j(x)$. Then, for almost every x, and except for a subsequence of zero logarithmic density,*

$$F_n(x) = o(|n(\lambda_n - \lambda_{n+1})|^{-\frac{1}{2}}).$$

In particular, if $\Sigma \mid c_n \mid^2/n^p < \infty$ for $p > 0$, $F_n(x) = o(n^{p/2})$. Starting from $\int \mid F_n \mid^2 = \sum_0^n \mid c_j \mid^2$, we have

$$\sum_0^\infty (\lambda_n - \lambda_{n+1}) \int \mid F_n \mid^2 = \sum_0^\infty (\lambda_n - \lambda_{n+1}) \sum_0^n \mid c_j \mid^2 = \sum_{j=0}^\infty \mid c_j \mid^2 \sum_{n \leqq j} (\lambda_n - \lambda_{n+1})$$

$= \Sigma \lambda_j \mid c_j \mid^2 < \infty$. Hence, for almost every x, the series $\Sigma (\lambda_n - \lambda_{n+1}) \mid F_n(x) \mid^2$ is convergent. Applying the theorem, our result follows.

Corollary 2. *Let $\Sigma \mid a_n \mid^2 < \infty$ and suppose that the series $\Sigma a_n \phi_n(x)$ is $(C, 1)$ summable to a function $g(x)$. Let $S_n(x) = \sum_0^n a_j \phi_j(x)$. Then, for almost every x and except for a subsequence of zero logarithmic density, $S_n(x)$ converges to $g(x)$* [4].

For, letting $\sigma_n(x) = (S_0 + \cdots + S_n)/(n+1)$, we have

$$(n+1)(S_n - \sigma_n) = \sum_0^n j a_j \phi_j = \Sigma c_j \phi_j,$$

with $\Sigma \mid c_n \mid^2/n^2 < \infty$. Applying the previous corollary in the form given above with $p = 2$, and $F_n(x) = (n+1)(S_n(x) - \sigma_n(x))$ we have $S_n - \sigma_n = o(1)$ for a. e. x and except for a subsequence of zero logarithmic density. Since $\sigma_n(x) \to g(x)$ for all x, the result follows.

Turning now to Theorem 2. 4, dealing with asymptotic density, we make the same choice of X and the measures as before, $X = \{1, 2, \cdots\}$, $K(n) = \{1, 2, \cdots, n\}$, $\mu_n(S) = \nu(S \cap K(n))/n$ where ν assigns mass 1 to each point. If $S(n)$ is the number of members of S in $K(n)$, then it is clear that $D(S) = \lim \mu_n(S) = \lim S(n)/n$, which is the usual formula for the asymptotic density of S. For convenience, we shall use "for almost all n" rather than "in density" in this instance. The corresponding notion of (μ)-summability is nothing more than $(C, 1)$ summability. Thus, Theorem 2. 4 takes the form:

Theorem 3. 2. *Let $\liminf S_n = L$ be finite, and $(C, 1)$-$\lim S_n = L$. Then $\{S_n\}$ converges to L for almost all n.*

Similarly, the corollary to this theorem becomes:

Corollary 1. *If* (i) *$\{S_n\}$ is bounded from below,* (ii) *for almost all n, $\liminf S_n \geqq L$, and* (iii) *$(C, 1) \lim S_n = L$, then $\{S_n\}$ converges to L for almost all n.*

As stated before, condition (i) is not superfluous. Let $a_n = 1 + (-1)^n$, and let b_n be $2k-1$ when $n = k^2$ and 0 when n is not a square. It is easily seen that $(C,1)\text{-lim}\, a_n = (C,1)\lim b_n = 1$. Setting $S_n = a_n - b_n$, we see that $(C,1)\text{-lim}\, S_n = 0$ while $S_n \geqq 0$ for almost all n. However, both 0 and 2 are limit points of $\{S_n\}$, each of density $\frac{1}{2}$, so that $\{S_n\}$ does not converge to 0 for almost all n.

This in turn leads to a result which is quite similar to, but independent of Theorem 3.1. This again improves on [3].

COROLLARY 2. *Let the series* Σc_n *converge, and suppose that* nc_n *is bounded from below, and that for almost all* n, $\liminf nc_n \geqq 0$. *Then,* nc_n *converges to zero for almost all* n.

Take

$$S_n = \sum_0^n c_j \text{ and } \sigma_n = (S_0 + S_1 + \cdots + S_n)/(n+1).$$

Since Σc_n converges, $\lim (S_n - \sigma_n) = 0$ so that $(C,1)\text{-lim}\, nc_n = 0$, and the conclusion follows.

COROLLARY 3. *Let* $A = \{a_1, a_2, \cdots\}$ *be a set of integers of unit density. Then,* $a_{n+1} = 1 + a_n$ *for almost all* n.

Set $b_n = a_{n+1} - a_n \geqq 1$. Since $\lim a_n/n = 1$, $(C,1)\text{-lim}\, b_n = 1$ and $\{b_n\}$ converges to 1 for almost all n. However, the b_n are integral so that $b_n = 1$ for almost all n.

COROLLARY 4. *Let* $f(z) = \Sigma a_n z^n$ *have unit radius of convergence, and let* $\Sigma |a_n|^2 < \infty$. *Let* $\lim f(z) = L$ *as* z *approaches* 1 *from inside the circle* $|z| < 1$. *Let* $S_n = \sum_0^n a_k$. *Then, for almost all* n, $\{S_n\}$ *converges to* L.

We may take $L = 0$ by altering a_0 only. Given $\epsilon > 0$ choose δ so that if $|z-1| < \delta$ and $|z| < 1$, then $|f(z)| < \epsilon$. Since $f(z)/(1-z) = \sum_0^\infty S_n z^n$, we have

$$\Sigma |S_n|^2 r^{2n} = \frac{1}{2\pi}\int_0^{2\pi} |f(z)/(z-1)|^2 \, d\theta \leqq \epsilon^2/(1-r^2) + M/\delta^2$$

where $M = \sum_0^\infty |a_n|^2 \geqq \frac{1}{2\pi}\int_0^{2\pi} |f(z)|^2 \, d\theta$ and $|z| = r < 1$. Hence,

$$\limsup\, (1-r^2)\Sigma |S_n|^2 r^{2n} \leqq \epsilon^2,$$

and since ϵ may be taken arbitrarily small, $\lim (1-r^2)\Sigma |S_n|^2 r^{2n}$ exists and is zero. Since Abel summability is an extension of $(C,1)$ summability, $(C,1)\text{-lim} |S_n|^2 = 0$, so that by the theorem, $\{S_n\}$ converges to 0 for almost all n.

A special implication of Theorem 3.2 is that a sequence $\{S_n\}$ which is $(C,1)$ summable to L, and for which L is either the upper or lower limit, is necessarily bounded for almost all n. It is easily seen that when L is unrestricted, $\{S_n\}$ is bounded off sets of arbitrarily small positive density. However, it is not in general possible to replace "arbitrarily small" by "zero."

THEOREM 3.3. *There is a sequence $\{S_n\}$ of integers with $(C,1)\text{-lim}\, S_n = 2$ which is unbounded on every set of upper density* 1.

Let A_k be the progression $(2^k N + 2^{k-1})_{N=0}^{N=\infty}$, for $k = 1, 2, \cdots$. These form a disjoint partition of the set of positive integers. Define the sequence by $S_n = k$ for all n in A_k; thus, $\{S_n\}$ begins $\langle 1, 2, 1, 3, \cdots \rangle$. As usual let $A_k(x)$ be the number of members of A_k not exceeding x. Then,

$$\sigma_m = 1/m \sum_1^m S_n = 1/m \sum_1^\infty k A_k(m) = \sum_1^{k_0} [m2^{-k} - 1/2](k/m)$$

where $k_0(m) = [\log(m/2)/\log 2] = O(\log m)$. Clearly, $\sigma_m \leqq \sum_1^\infty k2^{-k} = 2$, for every m. In the opposite direction,

$$\sigma_m \geqq \sum_1^{k_0} (m2^{-k} - 3/2) k/m \geqq \sum_1^{k_0} k2^{-k} - o(1)$$

so that as m increases, $\sigma_m \to 2$. Let M be a positive number and let B be a set of integers such that $|S_n| < M$ for all $n \in B$. B is necessarily disjoint from all the sets A_k when $k > M$. Hence,

$$\bar{D}(B) \leqq \sum_{k \leqq M} D(A_k) = \sum_{k \leqq M} 2^{-k} < 1$$

which proves the theorem.

The following simple result dealing with series of positive terms is a slight extension of a theorem of J. Arbault [1].

THEOREM 3.4. *Let $a_n > 0$ and suppose that $\Sigma\, 1/na_n < \infty$. Let p_n be an increasing sequence of positive numbers, with $np_n = O(P_n)$ where $P_n = p_1 + p_2 + \cdots + p_n \to +\infty$. Then, $(1/P_n)\sum_1^n p_k a_k \to +\infty$.*

Since $\Sigma\, 1/na_n$ converges, $\{1/a_n\}$ converges to 0 off a set of asymptotic density zero. Let D^* be the density defined by the measure ν^* which assigns mass p_n to the point n. Any set of asymptotic density zero is also of D^* density zero, so that $\{a_n\}$ converges to $+\infty$ on a set of unit D^* density. By Theorem 2.3, $(D^*)\text{-}\lim a_n = +\infty$ which is easily seen to be exactly the desired result. The p_n may be chosen as n^r where $r > 0$; the special choice $r = 1$ is the result obtained by Arbault.

Our next illustration is somewhat similar to the preceding. Consider the sequence-to-sequence transform defined by $\sigma_n = n^{-\rho} \sum_1^n a_k k^{\rho-1}$. When $\rho > 0$, this is convergence and boundedness preserving; however, when $\rho \leqq 0$, this may transform a convergent sequence into an unbounded sequence. We prove that if the transform is not bounded, it must be almost convergent to infinity; in particular, if the transform is bounded for almost all n, it is in fact bounded.

THEOREM 3.5. *Let* $\sigma_n = n^\lambda \sum_1^n a_k/k^{\lambda+1}$ *with* $\lambda \geqq 0$, *and let* $|a_n| \leqq 1$. *Then,* σ_n *is either bounded, or* $|\sigma_n|$ *converges to* $+\infty$ *on a set* S *of upper density* 1. *Moreover, if* S *contains a sequence* $\{p_n\}$ *with* p_{n+1}/p_n *bounded, then* $\lim |\sigma_n| = +\infty$.

If $k \geqq n$, then $|\sigma_k/k^\lambda| \leqq |\sigma_n/n^\lambda| + (n+1)^{-\lambda} + \cdots + k^{-\lambda} \leqq |\sigma_n/n^\lambda| + (1/\lambda)(n^{-\lambda} - k^{-\lambda})$. If $R > 1$ and $n \leqq k \leqq Rn$, then $|\sigma_k| \leqq |\sigma_n| R^\lambda + (R^\lambda - 1)/\lambda$. (When $\lambda = 0$, this last term is to be replaced by $\log R$). Given $M > 0$ let $M' = MR^\lambda + (R^\lambda - 1)/\lambda$. If $|\sigma_m| > M'$, then $|\sigma_n| > M$ for all n such that $m/R \leqq n \leqq m$. Let A be the set of integers n for which $|\sigma_n| < M$; assuming that $\{\sigma_n\}$ is unbounded, choose m so that $|\sigma_m| > M'$. The interval of integers from m/R to m is free of members of A, so that $A(m) = A(m/R)$. Dividing by m/R and letting m increase along the sequence for which σ_n is unbounded, we obtain $R\underline{D}(A) \leqq \bar{D}(A) \leqq 1$. Letting R increase, we have $\underline{D}(A) = 0$. As M increases, so do the sets A; appealing to the additivity theorem, the modified union of the sets A is a set B such that $\underline{D}(B) = 0$ while for any M, $|\sigma_n| > M$ for all sufficiently large integers n not in B. Thus, $|\sigma_n|$ converges to $+\infty$ on the complement S of B. S has upper density 1. If there is a sequence $\{p_n\}$ such that $p_{n+1}/p_n = O(1)$ and $|\sigma_{p_n}| \to +\infty$ then by the argument above, the intervals $[p_n/R, p_n]$ are disjoint from the set A for all sufficiently large n. If R is chosen so that $R > p_{n+1}/p_n$ for all n, then these intervals overlap, and the set A is finite. Since this is true for any choice of M, $|\sigma_n|$ converges to $+\infty$.

4. Complex values. Most of the theorems of Section 2 concerning the behavior of $f(x)$ as $x \to \infty$ in X go over to complex-valued functions; one is of sufficient interest to require separate treatment, namely Theorem 2.4, which asserts that a function which is summable to its limit superior or inferior must converge to that limit point on a set of unit density. We shall prove the complex form of this theorem, restricting ourselves for simplicity to the case of sequences and $(C, 1)$ summability. It is first necessary to choose a correct replacement for upper and lower limits for a complex sequence $\{S_n\}$. We find this in the notion of the core of a sequence, as introduced by Knopp. Let C_n be the closed convex hull of the infinite set $\{S_n, S_{n+1}, S_{n+2}, \cdots\}$ and let $C = \bigcap C_n$.

This closed convex set is called the core of $\{S_n\}$. It contains the convex hull of the set of limit points of $\{S_n\}$ and coincides with this set, if $\{S_n\}$ is bounded. In many theorems about summability of complex sequences, the core replaces the oscillation set of a real sequence. For example, if $\sigma_n = (S_1 + \cdots + S_n)/n$ the core of $\{\sigma_n\}$ is a subset of the core of $\{S_n\}$. We recall that a boundary point of a convex set is called extreme if it is not the mid point of two other points of the set. We introduce the term *outer limit point* for any of the extreme points of the core of a sequence $\{S_n\}$. It is easily seen that these are in fact limit points of $\{S_n\}$. When $\{S_n\}$ is real, the outer limit points are merely the upper and lower limits. For a bounded sequence, the core may then be described as the convex hull of its set of outer limit points.

THEOREM 4.1. *Let $\{S_n\}$ be a bounded complex sequence, and let it be Cesaro summable to one of its outer limit points, p. Then, $\{S_n\}$ converges to p for almost all n.*

We may assume that $p = 0$ and that the line through p which supports the core C is the real axis, and lies below C. Let $S_n = x_n + iy_n$; for any $\delta > 0$, all but a finite number of these lie within δ of C. We have $(C, 1)$-$\lim x_n = (C, 1)$-$\lim y_n = 0$ and $\liminf y_n \geqq 0$. By Theorem 3.2, $\{y_n\}$ converges to 0 for almost all n. Hence, for any $\epsilon > 0$, the set of n for which $y_n > \epsilon$ has zero density. Since 0 is an extreme point of C, the set C can touch the real axis on one side of the origin only. We suppose that the left side of C lies above this axis. Then, x_n is bounded from below, and for any $\delta > 0$, $x_n > -\delta$ for almost all n. Appealing to the stronger form, Corollary 1 of Theorem 3.2, $\{x_n\}$ converges to 0 for almost all n. Since the intersection of two sets of unit density has unit density, $\{S_n\}$ converges to 0 for almost all n.

We do not know if this theorem holds for unbounded sequences $\{S_n\}$. A slight modification of the above proof shows that the Theorem is valid in any case, if p is a *regular* extreme point of C, i. e. one for which there is a supporting line at p which contacts C nowhere else. In this case the Theorem also holds for a sequence $\{S_n\}$ of points in a Banach space, if in addition it is assumed that the closure of the set $\{S_1, S_2, \cdots\}$ is compact. The argument via real and imaginary parts of S_n may be replaced by the use of the functional F which supports C at p with $F(p) = 1$; one can immediately infer that $F(S_n)$ converges to 1 for almost all n. If a neighborhood of p is deleted, it is easily seen that the remaining S_n form a subsequence of zero density. This argument does not seem suited to the more precise theorem above. In the case of a general density, and a complex valued function $f(x)$, the core of f is to be taken as the intersection of the closed convex hulls of the sets $f(R_n)$.

UNIVERSITY OF WISCONSIN.

REFERENCES.

[1] J. Arbault, "Sur les séries à termes positifs," *Comptes Rendus*, vol. 222 (1946), pp. 217-219.

[2] R. C. Buck, "The measure-theoretic approach to density," *American Journal of Mathematics*, vol. 68 (1946), pp. 560-580.

[3] A. Denjoy, "Sur le convergence des séries Fourier," *Comptes Rendus*, vol. 207 (1938), pp. 210-213.

[4] A. Zygmund, "Sur le charactre de divergence des séries orthogonals," *Mathematica* (Cluj), vol. 9 (1935), pp. 86-88.

SQUARE SUMMATION AND LOCALIZATION OF DOUBLE TRIGONOMETRIC SERIES.*

By Victor L. Shapiro.

1. Introduction. Let $\sum a_M e^{iMX}$ be a double trigonometric series where a_M are arbitrary complex numbers and where $M = (m, n)$, $X = (x, y)$, $MX = mx + ny$, and $|M| = \max(|m|, |n|)$. The series will be said to be square convergent at a point X if the square partial sums of rank R

$$(1) \qquad S_R(X) = \sum_{|M| \leqq R} a_M e^{iMX}$$

converge to the finite value $L(X)$. The series will be said to be square summable (C, ρ), $\rho > 0$, to the sum $L(X)$ if the (C, ρ) square means of rank R,

$$(2) \qquad \sigma_R^{(\rho)}(X) = 2\rho R^{-2\rho} \int_0^R S_r(X)(R^2 - r^2)^{\rho-1} r\,dr$$

$$= \sum_{|M| \leqq R} a_M e^{iMX}(1 - |M|^2/R^2)^\rho = \sum_{r=0}^{[R]} (1 - r^2/R^2)^\rho \sum_{|M|=r} a_M,$$

converge to the finite value $L(X)$.

It is the purpose of this paper to study the localization theory of double trigonometric series for square summation. We shall use for this study the process of formal multiplication of series developed by Rajchman and Zygmund [2]. A comparison of the results obtained in this paper with those obtained by Berkovitz [1] for circular summation shows a decided difference between the two methods.

2. Definitions and notation. The notation in this paper will be for the most part vectorial, thus for example the index pair (m, n) will be designated by M, the capital letter of the first letter occurring, and $M + X$ will stand for $(m + x, n + y)$.

By $a_M = o[(|m| + 1)^\gamma(|n| + 1)^\eta]$ will be meant the following: Given an $\epsilon > 0$, there exists an $R(\epsilon)$ such that if $|M| > R(\epsilon)$, then

$$|a_M| < \epsilon(|m| + 1)^\gamma(|n| + 1)^\eta.$$

$a_M = O[(|m| + 1)^\gamma(|n| + 1)^\eta]$ will be defined in a similar manner.

* Received November 19, 1952; revised February 12, 1953.

Letting $T = \sum a_M e^{iMX}$, we set $\delta^j T/\delta x^j = \sum (im)^j a_M e^{iMX}$ ($\delta^0 T/\delta x^0$ will be interpreted as T); so that $\delta/\delta x$ is the symbol of partial differentiation $(= \partial/\partial x)$.

We shall designate the fundamental square $[(x,y); 0 \leqq x \leqq 2\pi, 0 \leqq y \leqq 2\pi]$ by Ω.

$f(X)$ will be said to be of class $C^{(k)}$ if all its partial derivatives up to and including those of order k exist and are continuous.

3. Formal multiplication. Let $T_1 = \sum a_M e^{iMX}$, $T_2 = \sum \alpha_M{}^{iMX}$ be two double trigonometric series. We define their formal product $T_1 T_2 = T_3$ to be the series $T_3 = \sum A_M e^{iMX}$, where $A_M = \sum\limits_P a_P \alpha_{M-P}$. The definition only makes sense when the A_M, called the formal product coefficients, are defined. In particular, if the a_M are bounded and $\sum\limits_M |\alpha_M| < \infty$, then A_M is defined for every M. We prove the following lemma regarding formal product coefficients.

Lemma 1. *Suppose that T_1 is a series with coefficients $a_M = o(|M|^\rho)$, $\rho \geqq 0$, and T_2 is a series with coefficients such that $\sum |\alpha_M|\,|M|^\rho < \infty$. Then A_M is defined for all M and A_M is $o(|M|^\rho)$.*

For $|A_M| \leqq (\sum\limits_{|P|<|M|/2} + \sum\limits_{|M|/2 \leqq |P| \leqq 2|M|} + \sum\limits_{|P|>2|M|}) |a_P \alpha_{M-P}|$. It is easy to see that the first and third sum are $o(1)$ and the second sum is $o(|M|^\rho)$, which proves the lemma.

In proving the basic theorems concerning formal products, certain lemmas will be required. We shall prove them first.

Lemma 2. *Let $\sum \alpha_M e^{iMX} = 0$ for X in a set E where*

$$\alpha_M = O[(|m|+1)^{-\theta}(|n|+1)^{-\theta}], \qquad \theta > 1.$$

Then there is a $K > 0$ such that

(i) $\sum\limits_{|M| \leqq R} |\alpha_M| \leqq K(R+1)^{1-\theta}$;

(ii) $|\sum\limits_{|M| \geqq R} \alpha_{M-P} e^{i(M-P)X}| \leqq K(|R - |P|| + 1)^{1-\theta}$ *uniformly for X in E;*

(iii) *if $p \geqq q \geqq 0$, then* $\sum\limits_{|M| \leqq R,\, n^2 > m^2} |\alpha_{M-P}| \leqq K(p-q+1)^{1-\theta}$;

(iv) *if $p \geqq q \geqq 0$, then, uniformly for X in E,*

$$|\sum_{|M| \leqq R,\, m^2 \geqq n^2} \alpha_{M-P} e^{i(M-P)X}| \leqq K(p-q+1)^{1-\theta} + K(|R-p|+1)^{1-\theta}.$$

To prove (i), we notice that

$$\sum_{|M|\geqq R} |\alpha_M| \leqq \sum_{i=0} \sum_{|M|=[R]+i} |\alpha_M|. \tag{3}$$

Observing also that the inner sum on the right of (3) is $O[(R+i)^{-\theta}]$, we have that the right side of (3) is $O(R^{1-\theta})$.

To prove (ii), we can assume, since $\sum |\alpha_M|$ is convergent, that

$$|P| \geqq R+1 \quad \text{or} \quad |P| \leqq R-1.$$

If the former holds then

$$\Big|\sum_{|M|\leqq R} \alpha_{M-P} e^{i(M-P)X}\Big| \leqq \sum_{|M|\geqq|P|-R} |\alpha_M|, \text{ while}$$

$$\Big|\sum_{|M|\leqq R} \alpha_{M-P} e^{i(M-P)X}\Big| \leqq \sum_{|M|>R-|P|} |\alpha_M|$$

if the latter holds, because the set $[M-P;\ |M|>R]$ lies outside of a square with center at the origin and side of length $R-|P|$. Applying (i) of this lemma to both cases completes the proof of (ii).

To prove (iii), we may suppose that $p>q+2$. Observing that the set $[M-P;\ m^2<n^2\leqq R^2]$ lies outside of a square with center at the origin and side of length $(p-q)/2$, we have that

$$\sum_{|M|\leqq R,\ n^2>m^2} |\alpha_{M-P}| \leqq \sum_{|M|\geqq|p-q|/2} |\alpha_M|.$$

Applying (i) of this lemma, we obtain the desired result.

(iv) is an immediate consequence of (ii) and (iii).

Lemma 3. *If* $a_M = o[(|m|+1)^{-\gamma}(|n|+1)^{-\eta}]$ *where* $0\leqq\gamma+\eta\leqq 1$, $0\leqq\gamma<1$, *and* $0\leqq\eta<1$, *then* $\sum_{|M|=i} |a_M| = o[i^{1-(\gamma+\eta)}]$.

This lemma follows in an obvious manner from the fact that

$$\sum_{n=-i}^{i} |a_{in}| = o\Big(i^{-\gamma}\sum_{n=-i}^{i}(|n|+1)^{-\eta}\Big) = o[i^{1-(\gamma+\eta)}].$$

We are now in a position to prove the following two theorems concerning formal products.

Theorem 1. *Let* T_1 *and* T_2 *be two double trigonometric series where* $T_1=\sum a_M e^{iMX}$ *and* $T_2=\sum \alpha_M e^{iMX}$ *with the following properties:*

(i) $a_M = o[(|m|+1)^{-\gamma}(|n|+1)^{-\eta}]$, $\gamma+\eta=1$; $\gamma, \delta\eta \geqq 0$;

(ii) $\alpha_M = O[(|m|+1)^{-3}(|n|+1)^{-3}]$;

(iii) $\sum \alpha_M e^{iMX} = 0$ *for* X *belonging to a plane set* E.

Then the formal product T_3 of T_1 and T_2 is uniformly square convergent to zero for X in E.

THEOREM 2. *Let T_1 and T_2 be two double trigonometric series where $T_1 = \sum a_M e^{iMX}$ and $T_2 = \sum \alpha_M e^{iMX}$ with the following properties:*

(i) $a_M = o[(|m|+1)^{-\gamma}(|n|+1)^{-\eta}]$, $0 \leqq \gamma + \eta < 1; \delta\eta \geqq 0$;

(ii) $\alpha_M = O[(|m|+1)^{-5}(|n|+1)^{-5}]$;

(iii) $\delta^j T_2/\delta x^j = 0$, $\delta^k T_2/\delta y^k = 0$, $j = 0,1,2$, $k = 0,1,2$, *for X belonging to a plane set E. Then the formal product T_3 of T_1 and T_2 is uniformly square summable $(C, 1-(\gamma+\eta))$ to zero for X in E.*

Remark 1. It will be apparent from the proofs of the theorems that (ii) in Theorem 1 can be replaced by $\alpha_M = O[(|m|+1)^{-(2+\epsilon)}(|n|+1)^{-(2+\epsilon)}]$ and that (ii) in Theorem 2 can be replaced by

$$\alpha_M = O[(|m|+1)^{-(4+\epsilon)}(|n|+1)^{-(4+\epsilon)}], \text{ where } \epsilon > 0.$$

That the formal product coefficients are defined in both theorems follows from Lemma 1. For simplicity of notation in the proofs of both theorems we shall suppose that 0 is in E and give the proof of the theorem only for this point. From the method of proof the uniformity of convergence or summability for all points in E will follow automatically.

Set

$$\sum_{|M|\leqq R} A_M(1-|M|^2/R^2)^\beta \tag{4}$$
$$= \Big(\sum_{|P|=0}^{[R]-1} \sum_{|P|=[R]}^{[R]+1} \sum_{|P|=[R]+2}^{[2R]} \sum_{|P|=[2R]+1}^{\infty}\Big) a_P \sum_{0\leqq|M|\leqq R} \alpha_{M-P}(1-|M|^2/R^2)^\beta$$
$$= A + B + C + D, \text{ where } \beta = 0 \text{ or } 1.$$

Then we see from (i) of Lemma 2 that under the conditions of either Theorem 1 or 2,

$$|D| \leqq \sum_{|P|>2R} |a_P| \sum_{|M|\leqq R} |\alpha_{M-P}| = o(R^{-1}). \tag{5}$$

Setting $\beta = 0$ in (4) and bearing (5) in mind, we observe that to prove Theorem 1 it only remains to show that A, B, and C are $o(1)$. But these facts follow easily from (i) and (ii) of Lemma 2 and Lemma 3, for by these lemmas, $|A|$, $|B|$, $|C|$ are majorized by

$$\sum_{i=0}^{[R]-1} O[(R-i)^{-2}] \sum_{|P|=i} |a_P|, \sum_{i=[R]}^{[R]+1} O(1) \sum_{|P|=i} |a_P|, \sum_{i=[R]+2}^{[2R]} O[(i-R)^{-2}] \sum_{|P|=i} |a_P|,$$

respectively, and each of these sums is $o(1)$.

To prove Theorem 2, we shall first show that under the conditions of this theorem the left side of (4) is $o(R^{-(\gamma+\eta)})$ when $\beta=1$. To do this, by a simple argument of symmetry, it is sufficient to consider a_P equal to zero except for points in the first octant, i. e., $a_P=0$ unless $p\geqq q\geqq 0$.

Applying the equality $m^2=p^2+2p(m-p)+(m-p)^2$, we observe that with $\beta=1$ the inner sum on the right side of (4) can be written as

$$
\begin{aligned}
(6)\qquad & R^{-2}\sum_{|M|\leqq R,\, n^2\leqq m^2}\alpha_{M-P}[R^2-p^2-2p(m-p)-(m-p)^2]^\beta \\
&+R^{-2}\sum_{|M|\leqq R,\, n^2>m^2}\alpha_{M-P}[R^2-p^2+p^2-q^2-2q(n-q)-(n-q)^2]^\beta,
\end{aligned}
$$

and consequently, from Lemmas 2 and 3, that $|A|$, $|B|$, $|C|$ are majorized by

$$
R^{-1}\sum_{p=0}^{[R]-1}o(p^{1-(\gamma+\eta)})(R-p)^{-2}+R^{-2}\sum_{p=0}^{[R]-1}o(p^{1-(\gamma+\eta)}),
$$

$$
R^{-1}\sum_{p=[R]}^{[R]+1}p^{1-(\gamma+\eta)},\qquad R^{-2}\sum_{p=[R]+2}^{[2R]}o(p^{2-(\gamma+\eta)})(p-R)^{-3}+o(p^{1-(\gamma+\eta)}),
$$

respectively; hence each of them is $o(R^{-(\gamma+\eta)})$.

From these last three facts and (5), we conclude that with $\beta=1$ the left side of (4) is $o(R^{-(\gamma+\eta)})$ and consequently that

$$
\sigma_R^{(1)}(0)=o(R^{-(\gamma+\eta)}). \tag{7}
$$

Theorem 2 is thus proved in the special case when $\gamma+\eta=0$. Let us suppose for the rest of the proof that $0<\gamma+\eta<1$.

Observing that for an integer j,

$$
S_j(0)=(2j+1)^{-1}[(j+1)^2\sigma_{j+1}^{(1)}(0)-j^2\sigma_j^{(1)}(0)]
$$

we conclude from (7) that

$$
S_R(0)=o(R^{1-(\gamma+\eta)}). \tag{8}
$$

By (2) and (8), in order to complete the proof of Theorem 2, it remains only to show that

$$
2(1-\gamma-\eta)R^{-2(1-\gamma-\eta)}\int_0^{R-1}S_r(0)(R^2-r^2)^{-(\gamma+\eta)}r\,dr=o(1). \tag{9}
$$

Observing, however, that by (7), $\int_0^R S_r(0)r\,dr=o(R^{2-(\gamma+\eta)})$, we conclude after integrating the integral in (9) by parts that this integral is $o(R^{2-2(\gamma+\eta)})$ and consequently that (9) holds, which gives us Theorem 2.

Remark 2. From the proofs of Theorems 1 and 2 it is evident that if the α_M were functions of X such that in Theorem 1,

$$\alpha_M(X) = O[(|m|+1)^{-3}(|n|+1)^{-3}]$$

uniformly in X and in Theorem 2,

$$\alpha_M(X) = O[(|m|+1)^{-5}(|n|+1)^{-5}]$$

uniformly in X, both theorems would still hold.

Unlike circular summation (see Berkovitz [1], p. 330), the formal product theorems using square summation cannot be extended to higher orders of summability. We show this with the following theorem:

THEOREM 3. *Given any $\epsilon > 0$ and any integer $k > 0$ there exists two double trigonometric series $T_1 = \sum a_M e^{iMX}$ and $T_2 = \sum \alpha_M e^{iMX}$ with the following properties:*

(i) $a_M = o(|M|^\epsilon)$;

(ii) $\alpha_M = 0$ *except for a finite number of* M;

(iii)) $\delta^{i+j}T_2/\delta x^i \delta y^j = 0$, $0 \leqq i+j \leqq k$ *for* $X = 0$,

such that the formal product T_3 of T_1 and T_2 is square summable (C, ρ), $\rho \geqq 0$, to infinity for $X = 0$.[1]

We choose for our series $T_1 = \sum_{m=2}^{\infty} b_m e^{imx}(e^{imy} + e^{i(m+1)y})/2$ where $b_j = j^\epsilon(\log j)^{-1}$, $j = 2, 3, 4, \cdots$ and $T_2 = (1 - e^{ix})^{k+1}$. Then

$$a_{mn} = b_m/2 \text{ if } n = m \text{ or } n = m+1 \text{ and } m \geqq 2; \text{ otherwise } a_{mn} = 0,$$

$$\alpha_{mn} = (-1)^m \binom{k+1}{m},\ 0 \leqq m \leqq k+1 \text{ and } n = 0; \text{ otherwise } \alpha_{mn} = 0.$$

Clearly conditions (i), (ii), and (iii) of the theorem are satisfied.

Now

$$A_M = \sum_P a_P \alpha_{M-P} = (b_{n-1}\alpha_{m-(n-1)\ 0} + b_n \alpha_{m-n\ 0})/2 \text{ if } n \geqq 2; \text{ otherwise } A_M = 0.$$

By (2), the theorem will be proved if it is shown that

$$S_R(0) = \sum_{|M| \leqq R} A_M \to \infty.$$

A short calculation shows that, for t an integer,

$$S_{t+k+1}(0) = \sum_{j=1}^{k} b_{t+k} \sum_{m=0}^{k+1-j} \alpha_{m\ 0} + b_{t+k+1}/2$$

$$= (t+k+1)^\epsilon[\log(t+k+1)]^{-1}K + o(t^{\epsilon-1}), \text{ where } K = \tfrac{1}{2} + \sum_{j=1}^{k} \sum_{m=0}^{k+1-j} \alpha_{m\ 0}.$$

[1] The author is indebted to the referee for suggesting a short proof to this theorem.

It is clear that $K \neq 0$, and consequently, $S_{t+k+1}(0) \to +\infty$ or $S_{t+k+1}(0) \to -\infty$ according to the sign of K.

We end this section on formal products with the following theorems:

THEOREM 4. *Let T_1 and T_2 be two double trigonometric series satisfying the conditions of Theorem 1, except that T_2 converges to a function $\lambda(X)$ which need not be zero. Then the series*

$$T_3 - \lambda T_1 = \sum A_M e^{iMX} - \lambda(X) \sum a_M e^{iMX}$$

is uniformly square convergent to zero for all X.

THEOREM 5. *Let T_1 and T_2 be two double trigonometric series where $T_1 = \sum a_M e^{iMX}$ and $T_2 = \sum \alpha_M e^{iMX}$ with the following properties:*

(i) $a_M = o[(|m|+1)^{-\gamma}(|n|+1)^{-\eta}]$, $0 \leqq \gamma + \eta < 1$; $\gamma, \eta, \geqq 0$;

(ii) $\alpha_M = O[(|m|+1)^{-5}(|n|+1)^{-5}]$;

(iii) $\delta^j T_2/\delta x^j = 0$, $\delta^k T_2/\delta^k = 0$, $j = 1, 2$, $k = 1, 2$ *for X in a set $E \subset \Omega$, the fundamental square. Designate the formal product of T_1 and T_2 by T_3 and the function to which T_2 converges on E by $\lambda(X)$. Then the series $T_3 - \lambda T_1$ is uniformly square summable $(C, 1 - (\gamma + \eta))$ to zero for X in E.*

We shall only give the proof for Theorem 4, Theorem 5 being proved in the same manner from Theorem 2.

Define T_2^* to be the trigonometric series $\sum \alpha_M^* e^{iMX}$ where $\alpha_M^* = \alpha_M$ if $|M| \neq 0$, $\alpha_0^* = \alpha_0 - \lambda(X)$. Since $\lambda(X)$ is a bounded function, $\alpha_M^*(X) = O[(|m|+1)^{-3}(|n|+1)^{-3}]$ uniformly for all X. Setting $A_M^* = \sum a_P \alpha_{M-P}^* = A_M - \lambda(X) a_M$, we have by Theorem 1 and Remark 2 that $\sum_{|M| \leqq R} A_M^* e^{iMX}$ converges to zero uniformly for all X.

4. Localization. We shall now apply the results of the formal product theorems to the problem of localization. This application will be prefaced, however, by a few remarks.

Remark 3. If $f(X)$ is in L and periodic of period 2π in each variable, we may associate to f its Fourier series. Thus $f \sim \sum c_M e^{iMX} = \mathfrak{S}[f]$. In what follows $\mathfrak{S}[f]$ will be used to denote the Fourier series of a function f. The square partial sums $S_R(X)$ of $\mathfrak{S}[f]$ are given by

$$S_R(X) = \sum_{M| \leqq R} c_M e^{iMX} = \pi^{-2} \int_0^{2\pi} \int_0^{2\pi} f(U) D_R(u-x) D_R(v-y)\, du\, dv,$$

where $D_R(t) = 1/2 \sum_{m=-[R]}^{[R]} e^{imt}$ is the Dirichlet kernel.

Remark 4. Let T be the double trigonometric series, $T = \sum c_M e^{iMX}$. An operator L^2, which will be called the double integral operator, is defined to act on T as follows:

$$L^2T = \frac{c_0 x^2 y^2}{4} + \frac{y^2}{2} \sum_{m=-\infty}^{\infty}{}' \, c_{m0}(im)^{-2} e^{imx} + \frac{x^2}{2} \sum_{n=-\infty}^{\infty}{}' \, c_{0n}(in)^{-2} e^{iny} + \sum_{M}{}'' \, c_M (mn)^{-2} e^{iMX}$$

where $'$ indicates the omission of the value zero, and $''$ indicates the omission of the values $(0, n)$ and $(m, 0)$. It is clear that L^2T converges uniformly to a function $F(X)$ if $c_M = o(1)$. We shall call $F(X)$ the function associated with T.

Remark 5. We shall now discuss another essential notion, that of a localizing function. Let $\mathfrak{R}$ be a closed domain contained in the interior of the fundamental square Ω. Let $\mathfrak{R}^0$ denote the interior of $\mathfrak{R}$, and $\mathfrak{R}'$ be another closed domain such that $\mathfrak{R}' \subset \mathfrak{R}^0$. A function $\lambda(X)$ which is continuous, of period 2π in each variable, has Fourier coefficients $O(|M|^{-\mu})$, μ being a sufficiently large positive integer, and such that $\lambda(X) = 0$ for X not in $\mathfrak{R} \pmod{2\pi}$ and $\lambda(X) = 1$ for X in $\mathfrak{R}' \pmod{2\pi}$, is called a localizing function for the domains $\mathfrak{R}$ and $\mathfrak{R}'$. That such a function can be constructed for given closed domains $\mathfrak{R}$ and $\mathfrak{R}'$ is a well known fact.

We now state the theorem from which we deduce our localization theorem.

THEOREM 6. *Let T_1 be a double trigonometric series with coefficients $a_M = o[(|m|+1)^{-\gamma}(|n|+1)^{-\eta}]$, $0 \leqq \gamma + \eta \leqq 1$, $0 \leqq \gamma < 1$, $0 \leqq \eta < 1$. Then the series L^2T_1 converges uniformly to a function $F(X)$. Furthermore let $\lambda(X)$ be a localizing function of class $C^{(14)}$ associated with the domains $\mathfrak{R}$ and $\mathfrak{R}'$ and whose Fourier coefficients are consequently*

$$\alpha_M = O[(|m|+1)^{-7}(|n|+1)^{-7}].$$

Then the difference

$$(10) \quad \Delta_R(X) = \sum_{|M| \leqq R} a_M e^{iMX} - \pi^{-2} \int_0^{2\pi}\int_0^{2\pi} F(U)\lambda(U) d^2D_R(u-x)/dx^2 d^2D_R(v-y)/dy^2 \, dudv$$

is uniformly summable $(C, 1-(\gamma+\eta))$ to zero for X in $\mathfrak{R}'$.

By Remark 4,

$$F(X) = a_0 x^2 y^2/4 + y^2/2 \sum_{m=-\infty}^{\infty}{}' \, a_{m0}(im)^{-2} e^{imx}$$
$$+ x^2/2 \sum_{n=-\infty}^{\infty}{}' \, a_{0n}(in)^{-2} e^{iny} + \sum_M{}'' a_M (mn)^{-2} e^{iMX}$$

$= F_4 + F_3 + F_2 + F_1$. We shall consider four different cases in the proof corresponding to F_1, F_2, F_3, and F_4.

For case 1, we assume that $a_M = 0$ if $m = 0$ or $n = 0$. Then F, in this case, is identical with F_1 and the right side of (10) is the square partial sum of rank R of the series $T_1 - \delta^4 \mathfrak{S}[F\lambda]/\delta x^2 \delta y^2$. By Theorem 4, the formal product $\mathfrak{S}[F]\mathfrak{S}[\lambda] = \mathfrak{S}[F\lambda]$. Let C_M designate the Fourier coefficients of this formal product and let d_M designate the coefficients of $\mathfrak{S}[\lambda]$. Then

$$m^2 n^2 C_M = \sum_P{}'' a_P d_{M-P} (pq)^{-2} [(m-p)^2 + 2p(m-p) + p^2]$$
$$\times [(n-q)^2 + 2q(n-q) + q^2]$$

and consequently $\Delta_R(X)$ is equal to the square partial sum of rank R of the series

$$(11) \quad T_1 - T_1 \mathfrak{S}[\lambda] - \sum_{\substack{k=0 \\ j+k \neq 4}}^{2} \sum_{j=0}^{2} \rho_{jk} \delta^{j+k} \mathfrak{S}[F]/\delta x^j \delta y^k \delta^{4-(j+k)} \mathfrak{S}[\lambda]/\delta x^{2-j} \delta y^{2-k}$$

where ρ_{jk} are constants. Now by Theorem 4 or Theorem 5 the partial sum of rank R of the series $T_1 - T_1 \mathfrak{S}[\lambda]$ and of each of the series in the sum in (11) is uniformly summable $(C, 1 - (\gamma + \eta))$ to zero for X in $\mathfrak{R}'$, and the proof of case 1 is complete.

For case 2, we assume that $a_0 = 0$ and $a_M = 0$ if $n \neq 0$. Then

$$T_1 = \sum_{m=-\infty}^{\infty}{}' \, a_{m0} e^{imx} \quad \text{and} \quad F(X) = G(x) y^2/2 = F_2(X),$$

where

$$G(x) = \sum_{m=-\infty}^{\infty}{}' \, a_{m0} (im)^{-2} e^{imx},$$

and furthermore, we observe that the right side of (10) is given by $T_1 - \delta^4 \mathfrak{S}[G\lambda y^2/2]/\delta x^2 \delta y^2$ and that the formal product $\mathfrak{S}[G]\mathfrak{S}[\lambda y^2/2] = \mathfrak{S}[G\lambda y^2/2]$. Proceeding as in case 1, we obtain that $\Delta_R(X)$ is the square partial sum of rank R of the series

$$(12) \quad T_1 - T_1 \mathfrak{S}[\delta^2(\lambda y^2/2)/\delta y^2]$$
$$- \sum_{i=0}^{i} \rho_i \delta^i \mathfrak{S}[G]/\delta x^i \delta^{2-i} \delta^2 \mathfrak{S}[\lambda y^2/2]/\delta x^{2-i} \delta y^2$$

where ρ_i are constants. Noticing that $\delta^2(\lambda y^2/2)/\delta y^2 = 1$ on $\mathfrak{R}'$, we conclude from Theorem 4 or Theorem 5 that the square partial sum of rank R of the series $T_1 - T_1\mathfrak{S}[\delta^2(\lambda y^2/2)/\delta y^2]$ and of each of the series in the sum in (12) is uniformly $(C, 1-(\gamma+\eta))$ summable to zero for X in $\mathfrak{R}'$. This completes the proof for case 2.

For case 3, we assume that $a_0 = 0$ and $a_M = 0$ if $n \neq 0$. Then $T_1(X) = \sum_{n=-\infty}^{+\infty} a_{0n}e^{inv}$ and $F(X) = F_3(X)$, and by an argument similar to case 2, we conclude that $\Delta_R(X)$ is uniformly summable $(C, 1-(\gamma+\eta))$ to zero for X in $\mathfrak{R}'$.

For case 4, we assume that $a_M = 0$ when $M \neq 0$. Then $T_1 = a_0$ and $F(\mathrm{X}) = a_0x^2y^2/4$, and consequently the right side of (10) is the square partial sum of rank R of the series $T_1 - \delta^4\mathfrak{S}[\lambda a_0x^2y^2/4]/\delta x^2\delta y^2$, which clearly converges to zero uniformly for X in $\mathfrak{R}'$.

Putting cases 1, 2, 3, and 4 together, we have the proof of the theorem, for T_1 can be considered a sum of four parts, one corresponding to each case.

It is at this point that the divergence in localization between square summation and circular summation can be seen. Theorem 6 cannot be extended to trigonometric series whose coefficients are $o(|M|^\epsilon)$, $\epsilon > 0$, by means of formal products because the key theorem in the proof is Theoerm 5, and Theorem 3 gives a direct contradiction to Theorem 5 for such coefficients. On the other hand, as Berkovitz [1] shows, no such difficulty exists in circular summation where the formal product theorems exist regardless of the order of the coefficients, and consequently, for circular summation, localization goes through.

Using Theorem 6, we can now state the main theorem for localization in square summation.

THEOREM 7. *If T and T' are two double trigonometric series with coefficients $o[(|m|+1)^{-\gamma}(|n|+1)^{-\eta}]$, $0 \leqq \gamma+\eta \leqq 1$, $0 \leqq \gamma < 1$, $0 \leqq \eta < 1$ and if the functions F and F' associated with T and T' are equal in a closed domain $\mathfrak{R}$ contained in the interior of the fundamental square Ω, then in every smaller closed domain $\mathfrak{R}'$ contained in $\mathfrak{R}^0$, the interior of $\mathfrak{R}$, the series $T - T'$ is uniformly square summable $(C, 1-(\gamma+\eta))$ to zero. The condition that $F - F'$ vanish in $\mathfrak{R}$ can be replaced by the condition that*

$$\int_0^{2\pi}\int_0^{2\pi} \lambda(U)[F(U) - F'(U)]\, d^2D_R(u-x)/dx^2\, d^2D_R(v-y)/dy^2\, du dv$$

be uniformly summable $(C, 1-(\gamma+\eta))$ to zero in every $\mathfrak{R}'$ where $\lambda(X)$ is a localizing function for $\mathfrak{R}$ and $\mathfrak{R}'$ of class $C^{(14)}$. In particular, this latter

result is alway true if $F - F'$ is a function of class $C^{(8)}$ in a domain $\mathfrak{R}_1{}^0$ containing $\mathfrak{R}$ and $\delta^4(F - F')/\delta x^2\delta y^2 = 0$ for X in $\mathfrak{R}^0$.

The last statement of the theorem follows from the fact that

$$\pi^{-2}\int_0^{2\pi}\int_0^{2\pi}\lambda(U)[F(U) - F'(U)]\, d^2D_R(u - x)/dx^2\, d^2D_R(v - y)/dy^2\, dudv$$

is the square partial sum of rank R of

$$\delta^4\mathfrak{S}[(F - F')\lambda]/\delta x^2\delta y^2 = \mathfrak{S}[\delta^4(F - F')\lambda/\delta x^2\delta y^2].$$

This latter Fourier series has coefficients $o[(|m| + 1)^{-2}(|n| + 1)^{-2}]$, and has square partial sums which converge to zero uniformly for X in $\mathfrak{R}'$.

The rest of the theorem follows immediately from Theorem 6.

It is to be noticed, in closing, that for the orders of Cesaro summability discussed, localization by squares except for the $(C, 1)$ case requires a weaker hypothesis then localization by circles since the condition $o[(|m| + 1)^{-\gamma}(|n| + 1)^{-\eta}]$ is weaker than $o[(m^2 + n^2)^{-(\gamma+\eta)/2}]$ for the values of γ and η considered in this paper.[2]

RUTGERS UNIVERSITY.

BIBLIOGRAPHY.

[1] L. D. Berkovitz, "Circular summation and localization of double trigonometric series," *Transactions of the American Mathematical Society*, vol. 70 (1951), pp. 323-344.

[2] A. Zygmund, *Trigonometrical series*, Warsaw (1935).

[2] The present paper is based on Part I of the author's doctoral dissertation submitted to the Mathematics Department of the University of Chicago. The author wishes to acknowledge his indebtedness to Professor Antoni Zygmund, under whose direction the dissertation was written.

AN IDEAL-THEORETIC CHARACTERIZATION OF THE RING OF ALL LINEAR TRANSFORMATIONS.*

By Kenneth G. Wolfson.

Introduction. It is well known that a simple ring that satisfies the minimum condition on right (left) ideals is isomorphic to the complete ring of linear transformations of a finite-dimensional vector space ([8], p. 67). This determination of the structure of such simple rings also serves as an abstract characterization of the ring of all linear transformations of a finite-dimensional space. In studying the structure theory of rings not restricted by the minimum condition, emphasis has been placed on the notion of a primitive ring ([10]), a ring which is isomorphic to a dense ring of linear transformations of an (in general) infinite-dimensional vector space. Although necessary and sufficient conditions have been given ([9], [10]) that an abstract ring be isomorphic to a dense ring, no characterization of the complete ring of linear transformations of an infinite-dimensional vector space is to be found in the literature. This ring of linear transformations also arises in projective geometry since the principal left ideals of this ring form a lattice which is projectively equivalent to the projective geometry determined by the lattice of subspaces of the underlying vector space ([4], p. 173 and [17], p. 6). In fact, if the rank of the underlying space is at least three, then the group of projectivities of the vector space is essentially the same as the group of automorphisms of the ring, the collineations corresponding to the inner automorphisms of the ring ([4], p. 187).

We shall determine necessary and sufficient conditions that an abstract ring be isomorphic to a ring $T(F, A)$, the set of all linear transformations of the vector space A over the (not necessarily commutative) field F. We also characterize the ring $T_\nu(F, A)$, which is the set of all linear transformations of A of rank $< \aleph_\nu$ over F (where $\aleph_\nu$ is some infinite cardinal). In particular, $T_0(F, A)$ the ring of all linear transformations of A which are of finite rank is characterized as a simple ring with minimal right ideals satisfying a certain annihilation condition on left ideals (Theorem 6.2).

* Received May 1, 1952.

If K is an abstract ring, a right annulet of K is an ideal which is the totality of right annihilators of a subset of K. Left annulets are defined similarly. Our main result (Theorem 7.5) may be stated as follows: *K is isomorphic to the ring $T(F, A)$ of all linear transformations of a vector space A over a field F if and only if*

(1) *K_0, the socle of K, is not a zero ring, and is contained in every non-zero two-sided ideal of K.*

(2) *If J is a left ideal of K which is annihilated on the right only by zero, then $J \geqq K_0$.*

(3) *The sum of two right (left) annulets is a right (left) annulet.*

(4) *K possesses an identity element.*

The main tool in the investigation is an extension of the Galois correspondence developed by Baer in [2], [4] between the annulets in the transformation rings and the subspaces of the underlying vector space. The main result in this direction is Theorem 2.8.

In examining the structure of the rings $T_\nu(F, A)$ an essential result (Theorem 5.1) is the fact that these rings are generated by their idempotents. Our characterization of $T(F, A)$ makes use of the characterization of the rings $T_\nu(F, A)$ although it is clear that the theorem may be proved directly. We prove in Theorem 8.2 that the structure of $T(F, A)$ is completely determined by its ideal theory.

1. Basic concepts. A *linear manifold* is a pair (F, A) consisting of an additive abelian group A and a (not necessarily commutative) field F, in which the field elements operate from the left on the elements of A in the obvious manner.

As usual, a *basis* of the linear manifold (F, A) is a maximal set of linearly independent elements of A. It is well known ([4], p. 14) that every linear manifold possesses a basis, and that the cardinal number of each basis is the same. The *rank* of the linear manifold (F, A) is the cardinal number of any basis of A.

A *linear submanifold* or *subspace* of A is a non-vacuous subset S of A satisfying $S + S = S$ and $FS = S$. To every subspace S of A there exists a subspace Q of A such that $S \cap Q = 0$ and $S + Q = A$ ([4], p. 12). We write $A = S \oplus Q$ in this case, and call Q a *complement* of S.

We recall that an endomorphism of an additive group G is a single-valued mapping of G into itself which preserves addition. A *linear transformation*

or *F-endomorphism* of the linear manifold (F, A) is an endomorphism of A which commutes with the elements of the field F. Thus a linear transformation σ of (F, A) satisfies $(a' + a'')\sigma = a'\sigma + a''\sigma$ for a', a'' in A and $(fa)\sigma = f(a\sigma)$ for a in A and f in F. The totality of linear transformations of the linear manifold (F, A) is denoted by $T(F, A)$. Defining addition and multiplication of linear transformations in the usual way, it is clear that $T = T(F, A)$ is a ring, the ring of all linear transformations of the group A over F. If σ is a linear transformation of (F, A) by its *rank* $r(\sigma)$ is meant the rank of the subspace $A\sigma$. If $\aleph_\nu$ is an infinite cardinal number, we denote by $T_\nu = T_\nu(F, A)$ the totality of elements σ in $T(F, A)$ which satisfy $r(\sigma) < \aleph_\nu$. For each ordinal ν, $T_\nu(F, A)$ is a two-sided ideal of $T(F, A)$. If J is a non-zero two-sided ideal of $T(F, A)$ then $J = T_\mu(F, A)$ for some ordinal μ ([4], p. 198). The same proof as given in [4] can be used to show that every non-zero two-sided ideal of the ring $T_\nu(F, A)$ is of the form $T_\mu(F, A)$ with $\mu \leqq \nu$.

We note that $T_0(F, A)$ consists of all linear transformations of A which are of finite rank, and that $T_0(F, A)$ is contained in every non-zero two-sided ideal of $T_\nu(F, A)$.

2. The Galois theory. In this section, the subspaces of (F, A) are related to certain classes of ideals in the rings of linear transformations. The methods follow those used by Baer in [2] and [4] for the particular ring $T(F, A)$. In some cases the proof may be the same but will be included for the sake of completeness.

Now let K be an arbitrary ring. If S is any subset of K then the totality $\mathfrak{R}(S)$ of elements x in K such that $Sx = 0$ is clearly a right ideal; and such a right ideal we term a *right annulet*. Likewise we denote by $\mathfrak{L}(S)$ the totality of elements y in K such that $yS = 0$. Clearly $\mathfrak{L}(S)$ is a left ideal; and such a left ideal we call a *left annulet*. We have the following:

LEMMA 2.1. *Let K be an arbitrary ring. Then every right annulet $J = \mathfrak{R}[\mathfrak{L}(J)]$ and every left annulet $H = \mathfrak{L}[\mathfrak{R}(H)]$.*

Proof. If H is a left annulet then $H = \mathfrak{L}(Q)$ for some subset Q of K. But $\mathfrak{L}[\mathfrak{R}(\mathfrak{L}(Q))] = \mathfrak{L}(Q)$ holds in any ring, and thus $H = \mathfrak{L}[\mathfrak{R}(H)]$. A similar proof holds for the right annulets.

Now let $E = E(F, A)$ denote any ring of linear transformations of the linear manifold (F, A). If S is a subset of A, then $R(S)$ is the totality of elements σ in E such that $S\sigma = 0$, and $L(S)$ is the totality of elements τ in E such that $A\tau \leqq S$. If S is actually a subspace, then $R(S)$ is a right

ideal and $L(S)$ is a left ideal, the annihilator of S, and the retraction on S respectively.

If J is a subset of E then $N(J)$ is the totality of elements x in A such that $xJ = 0$. By AJ is meant the set of elements aj for a in A and j in J. Although $N(J)$ is always a subspace of A, in general AJ will not be a subspace.

LEMMA 2.2. *Let $E(F, A)$ be any ring of linear transformations of A. Then we have:*

(i) $L(S)R(S) = 0$ *for every subset S of A.*

(ii) $L[N(J)] = \mathfrak{L}(J)$ *for every subset J of E.*

(iii) $R(AJ) = \mathfrak{R}(J)$ *for every subset J of E.*

Proof. (i) $AL(S) \leqq S$ by definition of $L(S)$, and $SR(S) = 0$ by definition of $R(S)$.

Hence $[AL(S)]R(S) = 0$, $A[L(S)R(S)] = 0$ and $L(S)R(S) = 0$.

(ii) is a consequence of the equivalence of the following statements:

$\sigma \, \varepsilon \, \mathfrak{L}(J)$, $\quad \sigma J = 0$, $\quad A\sigma J = 0$, $\quad A\sigma \leqq N(J)$, $\quad \sigma \, \varepsilon \, L[N(J)]$.

(iii) follows from the equivalence of:

$$\sigma \, \varepsilon \, \mathfrak{R}(J), \quad J\sigma = 0, \quad AJ\sigma = 0, \quad \sigma \, \varepsilon \, R(AJ).$$

COROLLARY 2.3. *If $E(F, A)$ is a ring of linear transformations and $AL(S) = S$ holds for a subspace S, then $R(S) = \mathfrak{R}[L(S)]$.*

Proof. By hypothesis $R(S) = R[AL(S)] = \mathfrak{R}[L(S)]$ by Lemma 2.2.

COROLLARY 2.4. *If $E(F, A)$ is a ring of linear transformations and $N[R(S)] = S$ holds for a subspace S, then $L(S) = \mathfrak{L}[R(S)]$.*

Proof. Since $S = N[R(S)]$ we have $L(S) = L[N(R(S))] = \mathfrak{L}[R(S)]$ by Lemma 2.2.

The ring $E(F, A)$ is called a dense ring of linear transformations if given any finite set of elements a_i $(i = 1, 2, \cdots, k)$ in A and linearly independent over F, and any set b_i $(i = 1, 2, \cdots, k)$ there exists a linear transformation e in $E(F, A)$ such that $a_ie = b_i$ for each i.

LEMMA 2.5. *Let $E(F, A)$ be a dense ring of linear transformations of A. Then $AL(S) = S$ for every subspace S of A if and only if $E(F, A)$ contains a dense ring of linear transformations of A of finite rank.*

Proof. Assume $E(F, A)$ contains a dense ring of linear transformations of finite rank. By definition of $L(S)$ it follows that $AL(S) \leqq S$. Now let $s \neq 0$ be in S. By assumption there exists a σ in $E(F, A)$ such that $s\sigma = s$ and $A\sigma$ has finite rank. Since $s\sigma = s$, $s \varepsilon A\sigma$. Let $s, m_1, \cdots, m_k$ be a basis of $A\sigma$. By density, there exists $\tau \varepsilon E$ such that $s\tau = s$ and $m_i\tau = 0$ for $i = 1, 2, \cdots, k$. Now $A\sigma\tau = Fs \leqq S$ and hence $\sigma\tau \varepsilon L(S)$. Then $s \varepsilon Fs = A\sigma\tau \leqq AL(S)$ and we have $S \leqq AL(S)$. Combined with the previous inequality, this yields $AL(S) = S$.

Now assume $AL(S) = S$ for every subspace S of A. Let S be of finite (positive) rank. If $L(S) = 0$ then $S = AL(S) = 0$ a contradiction. Hence there exists $e \neq 0$ in $L(S)$, and since $Ae \leqq S$, e has finite rank. Since the set of all transformations in $E(F, A)$ which are of finite rank is clearly a two-sided ideal, and since every non-zero two-sided ideal of a dense ring is again a dense ring ([10], p. 313) it follows that $E(F, A)$ contains a dense ring of linear transformations of finite rank.

LEMMA 2. 6. *If $E(F, A)$ is a dense ring of linear transformations, then $N[R(S)] = S$ holds for every subspace S of finite rank. The relation holds for every subspace if, and only if, $E(F, A)$ contains all the linear transformations of A which are of finite rank.*

Proof. Since $SR(S) = 0$, it follows that $S \leqq N[R(S)]$ for every subspace S. To prove $N[R(S)] \leqq S$ we need only show that if a is in A but not in S, then a doesn't belong to $N[R(S)]$. Let $A = S \oplus Fa \oplus U$. If S has finite rank, the density of $E(F, A)$ implies that there exists e in $E(F, A)$ satisfying $ae = a$ and $Se = 0$. If $E(F, A)$ contains all the linear transformations of finite rank, and S has infinite rank, the additional stipulation $Ue = 0$ assures the existence of the required e in $E(F, A)$. Since $Se = 0$ it follows that $e \varepsilon R(S)$. But $ae \neq 0$ implies $aR(S) \neq 0$ and hence $a \not\varepsilon N[R(S)]$ which completes the proof that $N[R(S)] = S$.

Now assume $N[R(S)] = S$ for every subspace S of A. Let S be an arbitrary hyperplane of A and assume, by way of contradiction that $R(S) = 0$. Then $S = N[R(S)] = N(0) = A$ a contradiction. Hence $R(S) \neq 0$ for hyperplanes S. Let σ be a linear transformation of finite rank n. Then we may write $A = \sum_{i=1}^{n} Fs_i \oplus N(\sigma)$, where the s_i are linearly independent. Let S_i be the hyperplane $\sum_{j \neq i} Fs_j \oplus N(\sigma)$. Then since $R(S_i) \neq 0$ there exists for $i = 1, 2, \cdots, n$, $\sigma_i \neq 0$ in $E(F, A)$ such that $S_i\sigma_i = 0$. If $s_i\sigma_i = 0$ it would follow $A\sigma_i = 0$ and hence $\sigma_i = 0$. Now by density of $E(F, A)$, there exists $\tau_i \varepsilon E(F, A)$ such that $(s_i\sigma_i)\tau_i = s_i\sigma$. Clearly $\sigma = \sum_{i=1}^{n} \sigma_i\tau_i$, and since σ_i, τ_i are

in the ring $E(F, A)$ for each i, it follows that $\sigma \varepsilon E(F, A)$ completing the proof.

LEMMA 2.7. *Let $E(F, A)$ contain a dense ring of linear transformations of finite rank. Then every left annulet $H = L(AH)$ and every right annulet $J = R[N(J)]$.*

Proof. Let H be a left annulet, so that $H = \mathfrak{L}(Q)$ for some subset Q of $E(F, A)$. But $\mathfrak{L}(Q) = L[N(Q)]$ by Lemma 2.2. Hence $AH = AL[N(Q)] = N(Q)$ by Lemma 2.5. Finally $H = L(AH)$.

If J is a right annlet then

$$J = \mathfrak{R}[\mathfrak{L}(J)] = R[A\mathfrak{L}(J)] = R[AL(N(J))] = R[N(J)].$$

We note that under the conditions imposed here on $E(F, A)$, AH is a subspace of A, whenever H is a left annulet.

A *projectivity* is a one-one mapping of one partially ordered set upon another partially ordered set which preserves the order relation.

A *duality* is a one-one mapping of one partially ordered set upon another partially ordered set which inverts the order relation.

If the partially ordered sets under consideration are lattices, it is clear that projectivities will also preserve cross-cuts and joins, while dualities will interchange cross-cuts and joins.

We now state the essential result of this section.

THEOREM 2.8. *Let $E(F, A)$ contain all those linear transformations of A which are of finite rank. Then*

(i) *The correspondences $R(S)$ and $N(J)$ are reciprocal dualities between the subspaces S of A and the right annulets J of E.*

(ii) *The correspondences $L(S)$ and AH are reciprocal projectivities between the subspaces S of A and the left annulets H of E.*

(iii) *The correspondences $\mathfrak{L}(J)$ and $\mathfrak{R}(H)$ are reciprocal dualities between the right annulets J and the left annulets H of E.*

Proof. (i) By Lemmas 2.6 and 2.7 and Corollary 2.3 we have $S = N[R(S)]$ and $J = R[N(J)]$, $R(S) = \mathfrak{R}[L(S)]$. Hence the correspondence is one-one between the class of all right annulets of E and the totality of subspaces of A. If $S \leqq Q$ are subspaces of A then $R(S) \geqq R(Q)$. If $U \leqq V$ are subsets of $E(F, A)$ then $N(U) \geqq N(V)$, and hence the correspondences are dualities.

(ii) Follows similarly from Lemmas 2.5 and 2.7 and Corollary 2.4.

(iii) Follows directly from Lemma 2.1.

Remark 1. If $E(F, A)$ contains all the linear transformations of A which are of finite rank, the totality of left (right) annulets of $E(F, A)$ forms a complete complemented modular lattice which is projectively equivalent (dual) to the projective geometry determined by the linear manifold (F, A).

Remark 2. The intersection of any number of left (right) annulets is again a left (right) annulet. Hence to every set of left (right) annulets there exists a smallest left (right) annulet containing all the annulets in the given set: the *join* of the annulets in the set. Since annulets are ideals this join will always contain the ideal-theoretical sum, but in general it will be larger.

Remark 3. If $E(F, A)$ contains only a dense ring of linear transformations of finite rank then the statements of Theorem 2.8 remain valid for the set of all finite-dimensional subspaces of A and a corresponding subclass of the set of all right (left) annulets of $E(F, A)$. For, a right ideal J is a right annulet if, and only if, $J = R(S)$ for a subspace S of A. Every left annulet has the form $L(S)$ for a subspace S of A, and if S is a subspace of finite rank then the ideal $L(S)$ is always a left annulet.

3. Primitive rings with minimal ideals.

An abstract ring K which is isomorphic to a dense ring of linear transformations is called a *primitive* ring.

The following is a collection of known results.

THEOREM 3.1. *Let $E(F, A)$ be a dense ring of linear transformations of A. Then E contains minimal right ideals if, and only if, E contains non-zero linear transformations of finite rank. In this case, the sum of all the minimal right ideals coincides with the sum of all the minimal left ideals, and with $E_0(F, A)$ the totality of linear transformations of A of finite rank which are contained in $E(F, A)$. The ring $E_0(F, A)$ is itself a dense ring of linear transformations of A which is a simple ring (not a zero ring) and is a two-sided ideal of $E(F, A)$ which is contained in every non-zero two-sided ideal of E.*

Proof. Every primitive ring has zero Jacobson radical ([10], p. 310). Since the radical contains all nilpotent ideals ([10], p. 304) it follows that a primitive ring contains no nilpotent ideals. Hence in particular E_0 is not a

zero ring. In any ring without nilpotent ideals the sum of all minimal right ideals coincides with the sum of all minimal left ideals ([11], p. 13). The remainder of the theorem is a restatement of Theorems 29 and 30 of [10] and the fact that any two-sided ideal of a dense ring is itself a dense ring.

LEMMA 3.2. *If K is a simple ring containing minimal right ideals, then K is semi-simple if, and only if, it is not a zero ring. In the event K is not a zero ring, it also contains minimal left ideals, and every right (left) ideal is the sum of minimal right (left) ideals.*

Proof. The sum of all minimal right ideals in any ring is a two-sided ideal ([7]). Since K is simple it is equal to the sum of all its minimal right ideals. If K is semi-simple then certainly it is not a zero-ring. Assume now $K^2 \neq 0$ but that $N \neq 0$ where N is the radical of K. Then since K is simple, $N = K$. Since every minimal right ideal is a left annihilator of the radical ([3], p. 565) it follows that $K^2 = KN = 0$ a contradiction. Hence $N = 0$. The fact that K contains minimal left ideals is due to Artin-Whaples ([1], p. 92). Hence K is also equal to the sum of its minimal left ideals. Thus, the lattice of right (left) ideals of K is a complete modular lattice in which the universal bound is the join of points. It follows then from a theorem of Birkhoff ([5], p. 129) that each element of the lattice is a join of points which completes the proof.

THEOREM 3.3. *Let $E(F, A)$ be a dense ring of linear transformations of A which contains minimal ideals, and let $E_0(F, A)$ be the two-sided ideal of linear transformations of finite rank which are contained in $E(F, A)$. Then*

(i) *Every right (left) ideal of $E_0(F, A)$ is a right (left) ideal of $E(F, A)$.*

(ii) *The minimal right (left) ideals of the rings E and E_0 are the same.*

(iii) *The minimal right (left) annulets are the same as the minimal right (left) ideals.*

(iv) *E_0 itself is a right (left) annulet if, and only if, $E_0 = E$.*

Proof. (i) The ring $E_0(F, A)$ contains minimal ideals since it contains non-zero linear transformations of finite rank. It is a simple (non-zero) ring by Theorem 3.1. By Lemma 3.2 every right (left) ideal is the sum of minimal right (left) ideals. Thus the proof of (i) will be complete if every minimal ideal of E_0 is an ideal of E. But in any ring K a minimal right ideal M satisfies either $M^2 = 0$ or $M = eM = eK$ where $e^2 = e$ is in M ([8],

p. 64). Since E_0 is semi-simple by Lemma 3.2, it contains no nilpotent ideals and $M = eE_0$ with $e^2 = e$ in M. Clearly $eE_0 \leqq eE$. But $e \varepsilon E_0$ implies $eE \leqq E_0$ since E_0 is a right ideal of E. Hence $eE = e(eE) \leqq eE_0$. Thus $M = eE_0 = eE$ and is clearly a right ideal of E. The same argument applies as well to the left ideals.

(ii) This is an immediate consequence of (i) and the fact that E_0 is the sum of all the minimal right (left) ideals of E.

(iii) If M is a minimal right ideal then $M = eE$ with e idempotent. Then $M = \mathfrak{R}[E(1-e)]$ where $E(1-e)$ is the set of elements $x - xe$ with x in E. Thus M is a right annulet, and since annulets are ideals, M is a minimal right annulet. Now, let M be a minimal right annulet. Now $ME_0 \leqq M \cap E_0$ and $ME_0 \neq 0$ since if $ME_0 = 0$ we would have $E_0{}^2 = 0$, a contradiction. Thus $M \cap E_0$ is a non-zero right ideal of E_0 and hence by Lemma 3.2 certainly contains a minimal right ideal M' which by previous remarks is a minimal right annulet. Since $M' \leqq M$ and both are minimal right annulets it follows $M = M'$ and M is a minimal right ideal. Since the same argument holds for left ideals and left annulets, this completes the proof of (iii).

(iv) If $E_0 = E$ then $E = \mathfrak{R}(0) = \mathfrak{L}(0)$ and is both a right and left annulet.

Now assume E_0 is a right annulet then $E_0 = R(S)$ for a subspace S. If $S = 0$, $E_0 = R(0) = E$ and we are finished. Assume therefore $S \neq 0$. Let $a \neq 0$ be in S then $aE_0 = 0$. But E_0 being a dense ring we must have $aE_0 = A$, a contradiction.

If E_0 is a left annulet, then $E_0 = L(S)$ for a subspace S of A. If $S = A$ then $E_0 = L(A) = E$. Assume therefore $S < A$. By density of E_0 we have $AE_0 = A$, but $AE_0 = AL(S) = S < A$ a contradiction. This completes the proof of the theorem.

Remark 1. Since every non-zero two-sided ideal I of $E(F, A)$ contains $E_0(F, A)$ it is clear that I is an annulet if, and only if, $I = E(F, A)$.

Remark 2. It is easy to see that an identity element of the ring E or E_0 if it exists acts as an identity transformation. Hence E_0 possesses an identity element if, and only if, the rank of A over F is finite.

Remark 3. If $E(F, A)$ is a dense ring and A has finite rank n, then $E(F, A)$ is the ring of all linear transformations of the linear manifold (F, A) or what is essentially the same thing, the ring of all n by n matrices with elements in the field F.

The situation as regards the relation of maximal ideals and maximal annulets is quite different than that of minimal ones. In fact we have the following:

THEOREM 3.4. *The following conditions on a ring K are equivalent:*

(1) *K is a primitive ring with an identity such that every maximal right ideal is an annulet.*

(2) *K is a primitive ring containing minimal ideals and an identity such that every maximal left ideal is an annulet.*

(3) *K is a primitive ring with minimal ideals and an identity such that the product of two right annulets is a right annulet.*

(4) *K is a primitive ring containing minimal ideals and an identity such that the product of two left annulets is a left annulet.*

(5) *K is a primitive ring with minimal ideals and an identity such that the sum of any set of right annulets is a right annulet.*

(6) *K is a primitive ring with minimal ideals and an identity such that the sum of any set of left annulets is a left annulet.*

(7) *K is a primitive ring in which every right and left ideal is an annulet.*

(8) *K is a primitive ring with an identity in which every left ideal is an annulet.*

(9) *K is (for some integer n) the ring of all n by n matrices over a field.*

Proof. It is clear that (9) implies all the other conditions since a total matrix ring is primitive, contains an identity and possesses minimal ideals since it satisfies the minimum condition on right (left) ideals. In fact, every ideal is generated by an idempotent and hence is an annulet ([8], p. 65). The fact that (7) implies (9) is a theorem of Kaplansky ([14], p. 694).

We shall show that each of the other conditions also implies (9).

Assume (1), then every maximal right ideal has the form $R(S)$ for S a subspace of rank 1. Also $N[R(S)] = S$ holds. Thus $L(S) = L[N(R(S))] = \mathfrak{L}[R(S)]$ by Lemma 2.2. If $L(S) = 0$ we have $\mathfrak{L}[R(S)] = 0$. But $R(S) = \mathfrak{R}[\mathfrak{L}(R(S))] = E(F, A)$ by Lemma 2.1, since $R(S)$ is an annulet. But this is impossible since $R(S)$ is a maximal ideal. Hence $L(S) \neq 0$ and

$E(F, A)$ contains non-zero linear transformations of finite rank and $E_0(F, A)$ is a dense ring. If $E_0 < E$ we may imbed (because of existence of an identity) E_0 in a maximal right ideal which is by assumption an annulet. From Remark 1 following Theorem 3.4, it follows $E_0 = E$ and from Remark 3 the conclusion follows.

Assume (2). It follows immediately that every maximal left ideal has the form $L(S)$ for S a hyperplane in A. Again if $E_0 < E$ we imbed E in a maximal left ideal $L(S)$ and the conclusion again follows.

Assume (3), and let S be a hyperplane in A. Now $R(0)R(S) = ER(S)$ is a two-sided ideal of $E \neq 0$, and a right annulet, whence the conclusion follows.

Assume (4), and let S be a subspace of rank 1, then $L(S)$ is an annulet and $L(S)L(A) = L(S)E$ is a two-sided ideal $\neq 0$ and is a left annulet from which the conclusion follows.

Assume (5) or (6). Since E_0 is the sum of all minimal right (left) annulets it follows that E_0 is an annulet from which the conclusion follows.

Assume (8), then K must be a simple ring. For assume $Q \neq 0$ is a two-sided ideal of K. Then Q is a dense ring ([10], Theorem 22). Since Q is a left ideal, it is an annulet. Let $Q = \mathfrak{L}(M) = L(S)$ where $S = N(M)$ (Lemma 2.2). If $S = A$, $Q = L(A) = K$. Hence assume $S < A$. Then $AL(S) \leqq S < A$. But since $L(S)$ is a dense ring $AL(S) = A$. This contradiction shows $Q = 0$ and hence K is simple. It follows from a theorem of Kaplansky ([15], p. 25) that K consists only of transformations of finite rank. Since K contains an identity the theorem follows.

This completes the proof of Theorem 3.4.

In the remainder of this section the ring $E(F, A)$ will be assumed to contain all the linear transformations of A which are of finite rank (that is, $E_0(F, A) = T_0(F, A)$ in our notation).

THEOREM 3.5. *Let $E(F, A)$ contain all the linear transformations of A which have finite rank. Then a left ideal J satisfies $\mathfrak{R}(J) = 0$ if, and only if, J contains all the linear transformations of A which are of finite rank.*

Proof. Assume $J \geqq T_0(F, A) = E_0(F, A)$. Then J is certainly a dense ring, and hence $AJ = A$. The fact that $\mathfrak{R}(J) = 0$ follows from the equivalence of the following statements:

$$\sigma \in \mathfrak{R}(J), \quad J\sigma = 0, \quad AJ\sigma = 0, \quad A\sigma = 0, \quad \sigma = 0.$$

Now assume $\mathfrak{R}(J) = 0$. Let $B \leqq A$ be the subspace spanned by the set of elements AJ. If $B < A$, there exists $\sigma \neq 0$ in $E(F, A)$ satisfying $B\sigma = 0$, $A\sigma \neq 0$. Hence $(AJ)\sigma \leqq B\sigma = 0$.

The following conditions however are equivalent:

$$(AJ)\sigma = 0, \quad A(J\sigma) = 0, \quad J\sigma = 0, \quad \sigma = 0, \text{ a contradiction.}$$

Hence $\{AJ\} = A$. Let $a \neq 0$ be in A. Then there must exist

$$a_1, a_2, \cdots, a_n \text{ in } A \text{ and } j_1, j_2, \cdots, j_n \text{ in } J \text{ such that } a = \sum_{i=1}^{n} a_i j_i.$$

There exist linear transformations of A of finite rank σ_i $(i = 1, 2, \cdots, n)$ in $E_0(F, A)$ such that $a\sigma_i = a_i$. Now $\sigma_i j_i \, \varepsilon \, J$ for $i = 1, 2, \cdots, n$ since J is a left ideal. Let $\tau = \sum_{i=1}^{n} \sigma_i j_i$, and $\tau \, \varepsilon \, J$. Thus $a\tau = a(\sum_{i=1}^{n} \sigma_i j_i) = \sum_{i=1}^{n} a_i j_i = a$. Now let μ be any linear transformation of rank 1, and let

$$A = Fa \oplus N(\mu) \text{ where } a\mu = b \neq 0.$$

By the above result, there exists α in J such that $b\alpha = b$. Since J is a left ideal $\mu\alpha \, \varepsilon \, J$. Thus $a(\mu\alpha) = b\alpha = b$, $N(\mu)\mu\alpha = 0$.

Hence $\mu = \mu\alpha \, \varepsilon \, J$. But as in the proof of Lemma 2.6 every finite transformation is the sum of transformations of rank 1, and therefore J contains all finite transformations.

Remark 1. If $E_0(F, A)$ is merely a dense ring then $J \geqq E_0$ implies $\mathfrak{R}(J) = 0$ but not conversely. For, we shall show that this condition assures that $E_0(F, A) = T_0(F, A)$ and there exist many examples of dense rings of linear transformations of finite rank which do not include all the linear transformations of finite rank.

THEOREM 3.6. *Let $E(F, A)$ contain all linear transformations of A which are of finite rank, and let M be a minimal right ideal of E. Then if I is the cross-cut of M and a left ideal H of $E(F, A)$, there exists a unique left annulet H^* such that $I = M \cap H^*$.*

Proof. We shall not give the proof in detail as it is essentially that given in [2] (Theorem 9.1) for the ring $T(F, A)$. It is shown there that AI is a subspace of A and $H^* = L(AI)$ is the required left annulet. We have already shown in Theorem 2.8 that the necessary properties of annulets hold in $E(F, A)$ since $E_0(F, A) = T_0(F, A)$. The modification necessary to take care of the fact that $E(F, A)$ need not possess an identity element is clear.

4. Ideal-theoretic properties of the annulets. In this section we shall examine the structure of annulets in the rings $T_\nu(F, A)$ and relate them to idempotents in the ring. In addition it will be shown that the rings $T_\nu(F, A)$ are generated by their idempotents.

Before proceeding we need the concept of *rank* for elements of a lattice. Let M be a lattice with zero element 0, and let x be an arbitrary element of $M \neq 0$. By *a chain between* 0 *and* x will be meant a well ordered (by the inclusion relation in the lattice) set of distinct elements of M which are bounded above by x, which includes 0, but not x. The *rank of* x is the least upper bound of the cardinal numbers of all chains between 0 and x, if $x \neq 0$. We define the rank of the zero element to be zero.

The totality of right (left) annulets of an arbitrary ring forms a lattice, and hence when we speak of the rank of a right (left) annulet we shall mean its lattice rank as defined above.

Let us consider the lattice of subspaces of a vector space V. If $x \neq 0$ is a subspace of V, consider any chain between 0 and x. If y, z are any elements of this chain and $y < z$, a basis of the subspace y may be extended to a basis of z. If we extend by one basis element at a time, and repeat this procedure for all such y and z, we may in this manner refine the given chain to a densest possible chain. The cardinal number of one of these chains, however, is clearly just the vector space rank of the subspace x. Hence we have shown that for the lattice of subspaces of a vector space the concept of lattice rank of a subspace coincides with its usual vector space rank.

Now let $E(F, A)$ be a dense ring of linear transformations containing all linear transformations of A, which are of finite rank. If J is a left annulet of E, there exists a unique subspace S such that $J = L(S)$. By the use of the preceding arguments and the projectivity of Theorem 2. 8 it follows that the rank of the left annulet $L(S)$ is just the vector space rank of S. Now let $H = R(S)$ be a right annulet of E. Since the zero element of the lattice of right annulets is $R(A)$, the preceding results imply that the rank of the right annulet $R(S)$ is the ordinary vector space rank of the quotient space A/S.

If J, J' are right (left) annulets of a ring K and $J \cap J' = 0, J \cup J' = K$, we shall say that J and J' are *complementary right* (*left*) *annulets* and either shall be called a *complement* of the other.

THEOREM 4. 1. (a) *If J is a right* (*left*) *annulet of $T_\nu(F, A)$ of rank $< \aleph_\nu$ and J' is a complementary right* (*left*) *annulet, then there exists an idempotent e in $T(F, A)$ such that*

$$J = eT_\nu, \quad J' = (1-e)T_\nu \quad (resp.\ J = T_\nu e, \quad J' = (1-e)T_\nu).$$

(b) *A right (left) annulet of* $T_\nu(F, A)$ *is generated by an idempotent if, and only if, it is of rank* $< \aleph_\nu$.

Proof. (a) Let H be a left annulet of rank $< \aleph_\nu$, and H' a complementary left annulet. Then it follows from Theorem 2.8 and our discussion of rank, that $H = L(U)$, $H' = L(W)$ where $A = U \oplus W$ and $r(U) < \aleph_\nu$. Define e as follows: $ue = u$ if $u \varepsilon U$, $We = 0$, so that $N(e) = W$. Then e is idempotent and $Ae = U$ tells us that $e \varepsilon T_\nu(F, A)$ since $r(e) < \aleph_\nu$. Since $Ae = U$ we have $e \varepsilon L(U) = H$. Since H is a left ideal $T_\nu e \leqq H$.

Now let $\tau \varepsilon H = L(U)$ so that $A\tau \leqq U$. Let a be in A. Then $a\tau = (a\tau)e$ since $a\tau \varepsilon U$ and $ue = u$ if $u \varepsilon U$. This implies $\tau = \tau e$ and thus $H \leqq T_\nu e$ which combined with previous inequality gives $H = T_\nu e$ with e idempotent. Now the following statements are equivalent:

$$x \varepsilon T_\nu(1-e), \quad xe = 0, \quad Axe = 0, \quad Ax \leqq Ne = W, \quad x \varepsilon L(W).$$

Hence $H' = T_\nu(1-e)$.

Let J be a right annulet of rank $< \aleph_\nu$, where J' is a complementary right annulet. Then it follows from Theorem 2.8 and our discussion of rank that $J = R(S)$, $J' = R(Q)$ where $A = S \oplus Q$ and $r(A/S) = r(Q) < \aleph_\nu$. Define the linear transformation e as follows: $Se = 0$, $qe = q$ if $q \varepsilon Q$, so that $S = N(e)$. Then e is idempotent and $Ae = Q$ so that $e \varepsilon T_\nu(F, A)$. Since $e \varepsilon R(S) = J$ which is a right ideal we have $eT_\nu \leqq R(S)$. Now let $f \varepsilon R(S)$ so that $Sf = N(e)f = 0$. If $a \varepsilon A$ then $a = ae + (a - ae)$ where $(a - ae)e = 0$ implies $a - ae \varepsilon S$ and therefore $(a - ae)f = 0$. Then $af = aef$ for each a in A implies $f = ef$ so that $R(S) \leqq eT_\nu$. Combined with the previous inequality, we have $J = eT_\nu$, with e idempotent. Now the following are equivalent:

$$x \varepsilon (1-e)T_\nu, \quad ex = 0, \quad Aex = 0, \quad Qx = 0, \quad x \varepsilon R(Q).$$

Hence $J' = (1-e)T_\nu$. This completes the proof of (a).

(b) By Theorem 2.8, every right (left) annulet of $T_\nu(F, A)$ possesses a complementary right (left) annulet. Hence we have already shown that every right (left) annulet of rank $< \aleph_\nu$ is generated by an idempotent element. Assume now H is a left annulet generated by an idempotent $H = L(U) = T_\nu e$ where $e^2 = e \neq 0$ is in T. If $a \neq 0$ and $a \varepsilon A$, we have $aT_\nu = A$ since T_ν is a dense ring. Then certainly $AT_\nu = A$. Now $U = AL(U)$ implies $U = AT_\nu e = Ae$. Since $e \varepsilon T_\nu$, $r(U) = r(Ae) < \aleph_\nu$ and by definition of rank, H is a left annulet of rank $< \aleph_\nu$. Now let J be a right annulet generated by an idempotent e so that $J = eT_\nu$. By Theorem 2.8, $J = R(S)$, S a subspace

of A. Since $N[R(S)] = S$ we have $S = N(eT_\nu)$. Clearly $N(e) = N(eT_\nu)$. We have $A = Ae \oplus Ne$ since e is idempotent, and thus $r(Ae) = r[A/N(e)]$. But since $e \in T_\nu(F, A)$, $r(Ae) < \aleph_\nu$ and hence $r[A/N(e)] = r(A/S) < \aleph_\nu$. By our discussion of rank for right annulets, we have that J is a right annulet of rank $< \aleph_\nu$. This completes the proof of (b).

Remark 1. It is clear from the proof that Theorem 4.1 is also valid for any ring $E(F, A)$ which contains the ring $T_\nu(F, A)$.

COROLLARY 4.2. *An ideal in $T(F, A)$ is an annulet if, and only if, it is generated by an idempotent.*

Proof. If $r(A) < \aleph_\nu$ then $T(F, A) = T_\sigma(F, A)$ for all ordinals $\sigma \geqq \nu$, hence every annulet is generated by an idempotent. We have previously remarked that in any ring, ideals which are generated by idempotents are certainly annulets. This completes the proof.

We note that Corollary 4.2 is proven in [4].

COROLLARY 4.3. *If J is a right (left) annulet of rank $< \aleph_\nu$ in $T_\nu(F, A)$ and J' is a complementary right (left) annulet, then T is the direct sum of the right (left) ideals J and J'.*

If J, J' are complementary right (left) annulets of $T(F, A)$, then T is the direct sum of the right (left) ideals J and J'.

A ring K is called *regular* if for every a in K, there exists an x in K such that $a = axa$. Such rings were introduced by von Neumann in [16] where existence of an identity was also assumed. However the following statements which are proved there are easily seen to be true without the existence of an identity element.

(1) In a regular ring, every principal right (left) ideal is generated by an idempotent element.

(2) In a regular ring, the sum of two principal right (left) ideals is a principal right (left) ideal.

THEOREM 4.4. *For each ordinal ν we have:*

(i) *The ring $T_\nu(F, A)$ is a regular ring.*

(ii) *The sum of two (and hence a finite number) of right (left) annulets of rank $< \aleph_\nu$ is a right (left) annulet of rank $< \aleph_\nu$.*

(iii) *Every right or left ideal which is finitely generated (and hence every principal ideal) is an annulet and is generated by an idempotent element.*

Proof. (i) The fact that $T(F, A)$ is a regular ring has been noted by Baer ([4], p. 179) and Johnson and Kiokemeister ([13], p. 407). Now let $e \varepsilon T_\nu$, then $e = efe$ for some f in $T(F, A)$. Then $e = (efe)fe = e(fef)e$ and fef is in T_ν since T_ν is a two-sided ideal of T. Hence T_ν is a regular ring. This last trick has been noted by Brown and McCoy ([6], p. 165).

(ii) An annulet of rank $< \aleph_\nu$ is generated by an idempotent (Theorem 4. 1) and is thus principal. Sum of two principal ideals in a regular ring is principal and generated by an idempotent and thus is an annulet of rank $< \aleph_\nu$ again by Theorem 4. 1.

(iii) A finitely generated ideal is a sum of finitely many principal ideals and is principal, since T_ν is a regular ring. Every principal ideal is generated by an idempotent and is therefore an annulet.

COROLLARY 4. 5. (Baer) *An ideal in $T(F, A)$ is an annulet if and only if it is finitely generated, and the sum of a finite number of left (right) annulets in $T(F, A)$ is a left (right) annulet.*

Proof. This is a consequence of Theorem 4. 4, and Corollary 4. 2.

Remark 1. If $T_\nu(F, A) \neq T(F, A)$ then the ring $T_\nu(F, A)$ possesses annulets which are not finitely generated. For T_ν itself is both a right and left annulet and if it were finitely generated it would be generated by an idempotent (Theorem 4. 4). This is impossible since such an idempotent would be an identity element for the ring $T_\nu(F, A)$.

Remark 2. In the rings $T_\nu(F, A)$ every ideal is the sum of annulets, since ideals are always sums of principal ideals. (Sums will, in general, be infinite).

5. The idempotents. We shall show in this section that the rings $T_\nu(F, A)$ are generated by their idempotents. The theorem we obtain is the following:

THEOREM 5. 1. *Let (F, A) be a linear manifold of rank at least two. If $E(F, A)$ is a ring of linear transformations which contains all idempotent transformations of A, of rank $< \aleph_\nu$, then $E(F, A)$ contains the ring $T_\nu(F, A)$. In particular if $E(F, A)$ contains all idempotent transformations of A, then $E(F, A) = T(F, A)$.*

The restriction to linear manifolds of rank at least two is essential, since it is clear that if (F, A) has rank one, the theorem fails to be true.

The proof will be given by proving three lemmas, each of which is a special case of the theorem.

LEMMA A. *Let* $A = \sum_\nu Fa_\nu \oplus N(\sigma)$ *where the* a_ν *are linearly independent, and the rank of* $N(\sigma)$ *is not less than that of* $\sum_\nu Fa_\nu$*, then* σ *belongs to the ring generated by those idempotents whose rank does not exceed that of* σ.

Proof. Let $a_\nu\sigma = b_\nu$. If the b_ν were dependent, then:

$$\sum_{i=1}^{m} f_ib_i = 0, \qquad \sum_{i=1}^{m} f_i(a_i\sigma) = 0, \qquad (\sum_{i=1}^{m} f_ia_i)\sigma = 0,$$

$$\sum_{i=1}^{m} f_ia_i \leqq \sum_\nu Fa_\nu \cap N(\sigma), \qquad \sum_{i=1}^{m} f_ia_i = 0$$

contradicting independence of the a_i. Since the mapping of a_ν onto b_ν is one to one it is clear that $\sum_\nu Fa_\nu$, and $\sum_\nu Fb_\nu$ have the same rank.

Define α as follows: $a_\nu\alpha = a_\nu$ for all ν, and $N(\sigma)\alpha = 0$. Then α is idempotent and has the same rank as σ.

We let $b_\nu = p_\nu + n_\nu$ where p_ν is in $\sum Fa_\nu$, n_ν is in $N(\sigma)$, and we shall map the a_ν firstly onto the p_ν. Since $r[N(\sigma)] \geqq r(\sum_\nu Fa_\nu)$ it follows that $A = \sum_\nu Fa_\nu \oplus \sum_\nu Fk_\nu \oplus W$ where $r(\sum Fk_\nu) = r(\sum Fa_\nu)$, the k_ν are linearly independent and each k_ν is in $N(\sigma)$, and $W \leqq N(\sigma)$. Since $\sum_\nu Fa_\nu \cap \sum_\nu Fk_\nu = 0$ it follows that $\sum_\nu F(a_\nu - k_\nu) \cap \sum_\nu Fk_\nu = 0$, and we may write

$$A = \sum_\nu F(a_\nu - k_\nu) \oplus \sum_\nu Fk_\nu \oplus W.$$

Define β as follows: $(a_\nu - k_\nu)\beta = 0$, $k_\nu\beta = k_\nu$, $W\beta = 0$. Then $a_\nu\beta = k_\nu$, β is idempotent, and rank $\beta =$ rank σ. Now since $\sum_\nu Fk_\nu \cap \sum_\nu Fp_\nu = 0$ we have also $\sum_\nu F(k_\nu - p_\nu) \cap \sum_\nu Fp_\nu = 0$. Hence we may write

$$A = \sum_\nu F(k_\nu - p_\nu) \oplus \sum_\nu Fp_\nu \oplus U$$

and define γ as follows: $(k_\nu - p_\nu)\gamma = 0$, $p_\nu\gamma = p_\nu$, $U\gamma = 0$. Then $k_\nu\gamma = p_\nu$, γ is idempotent, and rank $\gamma \leqq$ rank σ since $p_\nu \leqq \sum_\nu Fa_\nu$. Let $\alpha\beta\gamma = \tau$. Then $a_\nu\tau = p_\nu$, and $N(\sigma)\tau = 0$. We next must find a linear transformation ω which maps the a_ν onto the n_ν. Clearly $\sum_\nu Fa_\nu \cap \sum_\nu Fn_\nu = 0$ and therefore

also $\sum_\nu F(a_\nu - n_\nu) \cap \sum_\nu Fn_\nu = 0$. Hence we may write

$$A = \sum F(a_\nu - n_\nu) \oplus \sum Fn_\nu \oplus V$$

and define δ as follows: $(a_\nu - n_\nu)\delta = 0$, $n_\nu\delta = n_\nu$, $V\delta = 0$. Thus $a_\nu\delta = n_\nu$, δ is idempotent and rank $\delta \leqq$ rank σ since the n_ν need not be linearly independent while the b_ν are. Let $\omega = \alpha\delta$, then $a_\nu\omega = n_\nu$, and $N(\sigma)\omega = 0$. Now, $a(\tau + \omega) = p_\nu + n_\nu = b_\nu$, and $N(\sigma)(\tau + \omega) = 0$; therefore $\sigma = \tau + \omega$. This completes the proof, since none of the idempotents used had rank exceeding that of σ.

LEMMA B. *Let σ be a linear transformation of infinite rank. Then σ belongs to the ring of linear transformations generated by all those idempotent transformations whose rank does not exceed that of σ.*

Proof. Let $A = \sum Fa_\nu \oplus N(\sigma)$ where the a_ν are linearly independent and $a_\nu\sigma = b_\nu$. If $r[N(\sigma)] \geqq r(\sum Fa_\nu)$, the conclusion follows from the previous lemma. Hence we may assume $r[N(\sigma)] < r[\sum_\nu Fa_\nu]$. Since $\sum_\nu Fa_\nu$ has infinite rank it follows that $r(\sigma) = r[\sum_\nu Fa_\nu] = r(A)$. Therefore no restrictions are imposed on the ranks of idempotents used.

Now since we have infinite rank we may write $\sum Fa_\nu = \sum Fc_j \oplus \sum Fd_j$ where each subspace has the same rank, and the set consisting of all the c_j and d_j is merely the set of all the a_ν.

Since $r[\sum Fd_j \oplus N(\sigma)] = r(A) \geqq r(\sum_j Fc_j)$, there exists by the previous lemma a transformation α in the ring generated by the idempotents, which satisfies: $c_j\alpha = c_j$, $[\sum Fd_j \oplus N(\sigma)]\alpha = 0$.

In the same manner, since $r[\sum Fc_j \oplus N(\sigma)] \geqq r(\sum_j d_j)$ there exists an appropriate β satisfying: $d_j\beta = d_j$, and $[\sum_j Fc_j \oplus N(\sigma)]\beta = 0$. Clearly $a_\nu(\alpha + \beta) = b_\nu$ for each ν, $N(\sigma)(\alpha + \beta) = 0$ so that $\sigma = \alpha + \beta$, and the proof is complete.

The results of the two preceding lemmas are not directly applicable to linear transformations of finite rank, but by similar arguments we shall prove the following:

LEMMA C. *Let (F, A) be a linear manifold of rank at least two. Then the ring $T_0(F, A)$ coincides with the ring I generated by all the idempotents of finite rank.*

Proof. We show first the following:

(1) If u, v are independent elements of A, there exists a finite idempotent (of rank 1)σ such that $u\sigma = v$.

(2) If $A = Fa \oplus W$, there exists a finite idempotent of (rank 1)τ such that $a\tau = a$ and $W\tau = 0$.

To show (1) merely write $A = Fu \oplus Fv \oplus U$ and define σ by $u\sigma = v$, $v\sigma = v$, $U\sigma = 0$. Then σ is clearly idempotent of rank 1 and has required property. The idempotent τ needed in (2) is uniquely determined by the conditions imposed.

To show that we have all finite transformations in I, it suffices to show that I contains all transformations of rank 1. Thus let $A = Fa \oplus N(\alpha)$, $a\alpha = b$, where α is an arbitrary transformation of rank 1.

From (2) there exists τ such that $a\tau = a$ and $N(\alpha)\tau = 0$. If a, b are independent over F, then there exists by (1) σ such that $a\sigma = b$. Then $a(\tau\sigma) = b$, and $N(\alpha)\tau\sigma = 0$, so that $\alpha = \tau\sigma$. If a, b are dependent, then $b = fa$ where $f \neq 0$ is in F. Since $r(A) \geqq 2$ there exists d in $N(\alpha)$ such that a and d are linearly independent. From what we have already shown there exists an ω_1 in I such that: $a\omega_1 = d$, and $N(\alpha)\omega_1 = 0$. Now applying (1) there exists ω_2 in I such that $d\omega_2 = fa$, since d, fa are independent. Then $a\omega_1\omega_2 = fa$, and $N(\alpha)\omega_1\omega_2 = 0$. This completes the proof of the last lemma, since it is clear that $I \leqq T_0(F, A)$.

Theorem 5.1 now follows immediately from Lemmas B and C.

6. The ring $T_0(F, A)$. In this section we are interested in finding necessary and sufficient conditions that an abstract ring be isomorphic to the ring of all linear transformations which have finite rank.

We recall that a ring $P = P(A)$ of endomorphisms of the additive abelian group A is called *irreducible* if, for every a, not zero in A we have $aP = A$.

If a ring K contains minimal right ideals the *socle* of K is the sum of all its minimal right ideals. If K is without minimal right ideals then its socle is the zero ideal. The socle is always a two-sided ideal. (cf. Dieudonné [7]).

THEOREM 6.1. *Let K be an arbitrary ring. Then there exists a linear manifold (F, A) such that K is isomorphic to a ring $E(F, A)$ of linear transformations of A containing the ring $T_0(F, A)$ if, and only if,* (1) *and at least one of the conditions* (2), (3), *or* (4) *hold.*

(1) *The socle K_0 of K is not a zero ring and is contained in every non-zero two-sided ideal of K.*

(2) *If H is a left ideal of K and $\mathfrak{R}(H) = 0$ then $H \geqq K_0$.*

(3) *If H is a left ideal of K and $\mathfrak{R}(H) = 0$ then H contains a minimal right ideal.*

(4) *If M' is a minimal right ideal and J a left ideal of K then there exists a left annulet J^* such that $M' \cap J = M' \cap J^*$.*

Proof. Assume that K is isomorphic to $E(F, A)$ where $T_0(F, A) \leqq E(F, A) \leqq T(F, A)$. Then (1) holds by virtue of Theorem 3.1 and (2) is a consequence of Theorem 3.5.

Now if (1) and (2) hold then (1) and (3) are valid since (2) implies (3).

Assume now that (1) and (3) are valid. By (1) there exists a minimal right ideal M of K. Let I be the totality of elements x in K such that $Mx = 0$. Assume $I \neq 0$ so that $I \geqq K_0$ by (1) and thus $MK_0 = 0$. Now let I' be the totality of y in K satisfying $yK = 0$. Then we have $0 < M \leqq I'$ so that $I' \neq 0$. It follows from (1), since I' is a two-sided ideal, that $I' \geqq K_0$ or that $K_0{}^2 = 0$ which contradicts (1). Hence, we conclude that $I = 0$, that is $\mathfrak{R}(M) = 0$. In particular $M^2 \neq 0$ so that $M = eK = eM$ where $e^2 = e$ is in M.

For each k in K define an endomorphism σ_k of the additive group M by $x\sigma_k = xk$ for $x \in M$. The mapping of k onto σ_k constitutes a homomorphism of the ring K onto a ring $P(M)$ of endomorphisms of the group M. If σ_k is the zero endomorphism then $Mk = 0$ and $k = 0$ by preceding remarks. Hence K is actually isomorphic to the ring $P(M)$. We wish to show that $P(M)$ is an irreducible ring of endomorphisms. Let $y \neq 0$ be in M. Then $yK \leqq M$ since M is a right ideal. Since yK is a right ideal we have $yK = 0$ or $yK = M$ by the minimality of M. If $yK = 0$ then $yK_0 = 0$ and the preceding arguments lead again to the contradiction $K_0{}^2 = 0$. Hence $P(M)$ is an irreducible ring of endomorphisms. By Schur's Lemma ([8], p. 57) the totality of endomorphisms of M that commute with the elements of the irreducible ring $P(M)$ is a field F, and by Theorem 6 of [9] the ring of linear transformations $P(F, M)$ is a dense ring. Since $P(M)$ is a ring of right multiplications of M, where M is generated by an idempotent, it follows from [11] Lemma 1 that the commuting field F is isomorphic to $eKe = eMe$. It is easy to see that e is the identity of F and that e acts as an identity operator on the elements of M. Since $Me = (eMe)e = Fe$ the linear transformation σ_e of $P(F, M)$ is of rank 1 so that P contains transformations of finite rank. By Theorem 3.1, $P_0(F, M) = P_0(F, A)$ is a dense ring of

linear transformations of A which are of finite rank. We wish to show that $P_0(F,A)$ contains all linear transformations of A which have finite rank. As in the proof of Lemma 2.6 it is only necessary to show that $P_0(F,A)$ contains all linear transformations of rank 1, since every transformation of rank n is the sum of n transformations of rank 1. Since P is a dense ring, it is sufficient (as in Lemma 2.6) to find a non-zero element σ in P which annihilates an arbitrary hyperplane S of A. Assume by way of contradiction that S is a hyperplane of M, but that $R(S)=0$. By Corollary 2.3 $R(S)=\mathfrak{R}[L(S)]$. Hence $L(S)$ is a left ideal having right annihilator zero. It follows from (3) that $L(S)$ contains a minimal right ideal M'. By our construction of the ring $P(F,A)=P(F,M)$ we have $L(S)$ is the totality of x in K satisfying $Mx\leqq S$, so that $MM'\leqq S$. But $MM'\neq 0$ since $\mathfrak{R}(M)=0$, and therefore the minimality of M implies $MM'=M$. Hence $M\leqq S$ which contradicts the fact that S is a hyperplane in M. Hence $R(S)\neq 0$ and we have shown the sufficiency of (1) and (3).

If K is isomorphic to $E(F,A)$ where $T_0(F,A)\leqq E(F,A)\leqq T(F,A)$ then (4) holds by virtue of Theorem 3.6. Assume now (1) and (4). Then by virtue of (1), K is isomorphic to a ring $P(F,A)=P(F,M)$ where $P_0(F,M)$ is a dense ring of linear transformations of finite rank. Let $S<M$ be a hyperplane in M so that $S+S=S$ and $eMeS=MS=S$. Let J be the smallest left ideal of K containing S, so that J is the totality of elements of the form $\sum_{i=1}^{m}(k_is_i+n_is_i)$ where $k_i\,\varepsilon\, K$, $s_i\,\varepsilon\, S$ and n_i is an integer. Since $S<M$ and $S\leqq J$ we have $S\leqq M\cap J$. If $x\,\varepsilon\, M\cap J$, then $x=\sum(k_is_i+n_is_i)$ and $ex=x$ since $x\,\varepsilon\, M=eM=eK$. Hence

$$x=\sum(ek_is_i+n_ies_i)=\sum(ek_is_i+n_is_i)\,\varepsilon\, MS+S=S.$$

Hence we have $M\cap J\leqq S$, and combined with the previous inequality we have $M\cap J=S$. Now by (4), $S=M\cap J^*$ where J^* is a left annulet of K. If $J^*=K$ we have $S=M$ a contradiction. Hence $J^*<K$. Now assume, by way of contradiction, that $R(S)=0$. Since $S\leqq J^*<K$ we have $\mathfrak{R}(J^*)=\mathfrak{R}(K)=0$. By Lemma 2.1, $J^*=\mathfrak{L}[\mathfrak{R}(J^*)]=\mathfrak{L}[\mathfrak{R}(K)]=K$ and this contradicts $J^*<K$. Hence $R(S)\neq 0$ and again it follows that $P(eMe,M)$ contains the ring $T_0(eMe,M)$. This completes the proof of Theorem 6.1.

Let $\sigma>\nu$, then if $T_\sigma\neq T_\nu$ the ring $E=T_\sigma/T_\nu$ is a primitive ring containing no minimal ideals ([15], p. 18). Hence E contains no non-zero transformations of finite rank. Since $E_0=0$ the ring satisfies (2) but fails to satisfy all of (1) since the socle is a zero ring.

Since the condition (2) seems more desirable than (3) or (4) we shall use this condition in our future characterizations.

We are now in a position to characterize $T_0(F, A)$ itself.

THEOREM 6.2. *Let K be an arbitrary ring. Then there exists a linear manifold (F, A) such that K is isomorphic to the ring $T_0(F, A)$ of all linear transformations of A which are of finite rank if and only if*

(1) *K is a simple ring (not a zero ring) containing minimal right ideals.*

(2) *If H is a left ideal of K and $\Re(H) = 0$ then $H = K$.*

Proof. Since K is simple and possesses minimal right ideals we have $K = K_0$ its socle (since the socle is always a two-sided ideal). Hence K satisfies conditions (1) and (2) of Theorem 6.1 and therefore K is isomorphic to $E(F, A)$ where $T_0(F, A) \leqq E(F, A) \leqq T(F, A)$. But $E(F, A)$ in this case is simple, and since $T_0(F, A)$ is always a two-sided ideal we have $E(F, A) = T_0(F, A)$.

Now $T_0(F, A)$ satisfies (2) by Theorem 3.5 and (1) by Theorem 1 of [9]. This completes the proof.

An example of a dense ring of linear transformations of finite rank $P(F, A)$ which satisfies $P(F, A) = P_0(F, A) < T_0(F, A)$ is easily constructed as follows. Let (F, A) be a linear manifold with a countable basis, and let $P(F, A) = P_0(F, A)$ denote the ring of linear transformations which consists of only those transformations which annihilate all but a finite number of the basis elements. Then $P(F, A)$ is certainly a dense ring of transformations of finite rank, but does not, for example, contain the linear transformation of finite rank which maps every basis element into a fixed basis element b. Hence, $P(F, A)$ satisfies (1) but not (2) of the theorem since $P(F, A) < T_0(F, A)$.

7. The rings $T_\nu(F, A)$.

In this section we shall characterize the rings $T_\nu(F, A)$ and from this derive a characterization of $T(F, A)$.

As ideals of $T(F, A)$ the rings $T_\nu(F, A)$ are easily characterized abstractly as follows:

THEOREM 7.1. *The ideal $T_\nu(F, A)$ of $T(F, A)$ is the sum of all right annulets of rank $< \aleph_\nu$, and the sum of all left annulets of rank $< \aleph_\nu$.* (For the definition of rank of annulets see the beginning of Section **4**).

Proof. Let $\sigma = \sum_{i=1}^{k} \sigma_i$ where σ_i belongs to a right annulet of rank $< \aleph_\nu$,

that is $\sigma_i \varepsilon R(S_i)$ where $r(A/S_i) < \aleph_\nu$. Now, $A\sigma \leqq \sum_{i=1}^{k} A\sigma_i$, hence $r(A\sigma) \leqq \sum_{i=1}^{k} r(A\sigma_i)$. But $r(A\sigma_i) \leqq r(A/S_i) < \aleph_\nu$, and hence $r(A\sigma) < \aleph_\nu$. Then σ is a linear transformation of rank $< \aleph_\nu$ and $\sigma \varepsilon T_\nu(F, A)$.

Now let $\tau \varepsilon T_\nu(F, A)$ where $A = S \oplus N(\tau)$ and thus $r(A\tau) = r(S\tau) = r(S) < \aleph_\nu$. Then $\tau \varepsilon R[N(\tau)]$ a right annulet of rank $< \aleph_\nu$ and hence certainly belongs to a sum of such right annulets. This completes the proof of the assertion concerning right annulets.

Now consider left annulets, let $\sigma = \sum_{i=1}^{k} \sigma_i$ where $\sigma_i \varepsilon L(S_i)$ and $r(S_i) < \aleph_\nu$. Then $A\sigma_i \leqq S_i$ by definition of $L(S_i)$,

$$A\sigma \leqq \sum_{i=1}^{k} A\sigma_i \leqq \sum_{i=1}^{k} S_i = S$$

where $r(S) < \aleph_\nu$. Thus the rank of σ is less than $\aleph_\nu$ and $\sigma \varepsilon T_\nu(F, A)$. Again if $\tau \varepsilon T_\nu(F, A)$ then $A\tau = S$ where $r(S) < \aleph_\nu$. Thus $\tau \varepsilon L(S)$ a left annulet of rank $< \aleph_\nu$. This completes the proof.

Remark 1. For the special case of $\nu = 0$, we have that $T_0(F, A)$ is both the sum of all the minimal right ideals, and the sum of all the minimal left ideals of $T(F, A)$, which we previously had shown.

If $E(F, A)$ is a ring of linear transformations then the totality of linear transformations of rank $< \aleph_\nu$ (for any infinite cardinal) contained in $E(F, A)$ is a two-sided ideal. The arguments used in the preceding theorem remain valid in $E(F, A)$ if $E_0(F, A) = T_0(F, A)$ by Theorem 2. 8. Hence we have also proved the following:

COROLLARY 7. 2. *Let $E(F, A)$ be a dense ring of linear transformations containing all linear transformations of A which are of finite rank, then the two-sided ideal $E_\nu(F, A)$ of all transformations of rank $< \aleph_\nu$ in $E(F, A)$ coincides with the sum of all right annulets of rank $< \aleph_\nu$ and with the sum of all left annulets of rank $< \aleph_\nu$.*

THEOREM 7. 3. *Let K be an arbitrary ring. Then there exists a linear manifold (F, A) such that K is isomorphic to a ring $E(F, A)$ of linear transformations of A containing the ring $T_\nu(F, A)$ if, and only if, K satisfies conditions* (1), (2), (3) *below.*

(1) *The socle K_0 of K is not a zero ring and is contained in every non-zero two-sided ideal of K.*

(2) *If H is a left ideal of K and $\mathfrak{R}(H) = 0$ then $H \geqq K_0$.*

(3) *If J is a left (right) annulet of K of rank $< \aleph_\nu$ and J' is a left (right) annulet of K complementary to J, then there exists an idempotent e in J such that $J = Ke$, $J' = K(1-e)$ (resp. $J = eK$, $J' = (1-e)K$).*

We note again that (3) is not meant to imply the existence of an identity element in K. As far as sufficiency is concerned either the statement regarding right or left annulets is enough. Of course, both statements will be shown to be necessary.

Proof. The necessity of (1) and (2) follows from Theorem 6. 1. Condition (3) is necessary by virtue of Theorem 4. 1 and Remark 1 following that theorem.

Now assume K satisfies (1), (2), and (3). By Theorem 6. 1, K is isomorphic to a ring of linear transformations $E(F, A)$ which contains $T_0(F, A)$. If the rank of A over F is one, then since $E(F, A)$ is a dense ring, it is already the ring of all linear transformations of A of rank $< \aleph_\nu$ for every ordinal ν. Hence without loss of generality we may assume $r(A) \geqq 2$. In order to show that $E(F, A) \geqq T_\nu(F, A)$ it is, by virtue of Theorem 5. 1, only necessary to show that E contains all idempotent linear transformations of rank $< \aleph_\nu$. If e is any idempotent in $T_\nu(F, A)$, we may write $A = Ae \oplus N(e)$ where e is the identity on the subspace Ae of rank $< \aleph_\nu$, and an annihilator of the subspace $N(e)$. Hence to prove $E \geqq T_\nu$ we must show that to each decomposition $A = S \oplus Q$ where $r(S) < \aleph_\nu$, there exists a transformation in $E(F, A)$ which is the identity on S and annihilates Q. To this end, assume $A = S \oplus Q$ where $r(S) < \aleph_\nu$ and assume (3) for left annulets. Then by Theorem 2. 8, $E = L(S) \cup L(Q)$, and $0 = L(S) \cap L(Q)$, where $L(S)$ is a left annulet of rank $< \aleph_\nu$. Then $J = L(S)$ and $J' = L(Q)$ are complementary left annulets satisfying the hypothesis of (3). Hence by (3) $L(S) = Ee$ and $L(Q) = E(1-e)$, where $e^2 = e \neq 0$. But $S = AL(S) = AEe$ implies that the idempotent e is an identity on S. Now $Q = AL(Q) = AE(1-e)$ implies that $Qe = 0$ since $(1-e)e = 0$.

Similarly if we assume (3) for right annulets we have $E = R(S) \cup R(Q)$, and $0 = R(S) \cap R(Q)$, where $R(Q)$ is a right annulet of rank $< \aleph_\nu$, so that by (3) there exists an idempotent e in $R(Q)$ satisfying $R(Q) = eE$ and $R(S) = (1-e)E$. Since $e \varepsilon R(Q)$ we have $Qe = 0$. Now $SR(S) = 0$ implies $S[(1-e)E] = 0$ or $[S(1-e)]E = 0$. Hence $S(1-e) = 0$, since E annihilates only the zero subspace. Hence if $s \varepsilon S$, $s - se = 0$ or $s = se$, and e is the identity on S. Thus $E(F, A) \geqq T_\nu(F, A)$ which completes the proof.

THEOREM 7.4. *Let K be a ring of linear transformations which contains the ring $T_\nu(F, A)$. Then the following conditions are equivalent:*

(i) $K = T_\nu(F, A)$.

(ii) *K is equal to the sum of all its left (right) annulets of rank $< \aleph_\nu$.*

(iii) *The proper two-sided ideals of K form a well ordered set of order type ν.*

Proof. Assume (i). The ring $T_\nu(F, A)$ satisfies (ii) by virtue of Corollary 7.2 applied to the ring $T_\nu(F, A)$. Now the ordinal number ν represents the order type of the well ordered set of infinite cardinal numbers preceding $\aleph_\nu$. Hence, by the definition of $T_\nu(F, A)$, and the fact that all the non-zero two-sided ideals of T_ν have the form T_σ with $\sigma \leqq \nu$, it follows that ν is also the ordinal number of the well ordered set of proper two-sided ideals of $T_\nu(F, A)$. Hence (iii) is true.

Now assume $K \geqq T_\nu(F, A)$ and (ii) holds. Then Corollary 7.2 implies that K contains no linear transformations of rank $\geqq \aleph_\nu$ or $K = T_\nu(F, A)$. Clearly the assumption that $K > T_\nu(F, A)$ also violates (iii). Hence (iii) also implies that $K = T_\nu(F, A)$. This completes the proof of Theorem 7.4.

Remark 1. It is a consequence of the last theorems that the conditions (1), (2), (3) of Theorem 7.3 and (ii) or (iii) of Theorem 7.4 characterize the ring $T_\nu(F, A)$.

THEOREM 7.5. (MAIN THEOREM). *Let K be an arbitrary ring. Then there exists a linear manifold (F, A) such that K is isomorphic to the ring $T(F, A)$ of all linear transformations of A if and only if*

(1) *The socle K_0 of K is not a zero-ring and is contained in every non-zero two-sided ideal of K.*

(2) *If H is a left ideal of K and $\mathfrak{R}(H) = 0$ then $H \geqq K_0$.*

(3) *The sum of two right (left) annulets is a right (left) annulet.*

(4) *K possesses an identity element.*

Proof. Let J, J' be complementary right (left) annulets so that $J \cap J' = 0$ and $J \cup J' = K$. Since by (3) sums of right (left) annulets are right (left) annulets, it follows that $K = J \oplus J'$. Since K possesses an identity element it follows by a lemma of von Neumann ([16], p. 708) that there exists an idempotent e in K such that $J = eK$, $J' = (1 - e)K$ [resp. $J = Ke$, $J' = K(1 - e)$]. Hence K satisfies (1), (2), and (3) of Theorem

7. 3 for every ordinal ν. Hence by Theorem 7. 3, K is isomorphic to $E(F, A)$ where $T_\nu(F, A) \leqq E(F, A) \leqq T(F, A)$ for every ν and thus $E(F, A) = T(F, A)$.

It is clear that $T(F, A)$ possesses an identity element, and satisfies (1) and (2) by Theorem 7. 3. The fact that $T(F, A)$ satisfies (3) is a consequence of Corollary 4. 5. This completes the proof.

Remark 1. It is clear that the full strength of (3) was not needed. The following is sufficient:

(3′) If J and J' are complementary right (left) annulets of K, then K is the direct sum of the right (left) ideals J and J'.

8. Uniqueness theorems. A *semi-linear transformation* of a linear manifold (F, A) upon a linear manifold (G, B) is a pair $\sigma = (\sigma', \sigma'')$ consisting of an isomorphism σ' of the additive group A upon the additive group B, and an isomorphism σ'' of the field F upon the field G subject to the condition $(fa)^{\sigma'} = f^{\sigma''}a^{\sigma'}$ for f in F and a in A. We note that σ is a linear transformation if $F = G$ and $\sigma'' = 1$. There will be no confusion if in the future we use the same symbol σ both for the isomorphism σ' of A upon B and the isomorphism σ'' of F upon G.

Since $T_\nu(F, A)$ is for every ν, a dense ring containing minimal right ideals, a direct application of a theorem of Jacobson ([10], p. 318) yields the following result:

THEOREM 8. 1. *If α is an isomorphism of $T_\nu(F, A)$ upon $T_\nu(G, B)$, then there exists a semi-linear transformation σ of (F, A) upon (G, B) such that $t^\alpha = \sigma^{-1}t\sigma$ for every t belonging to $T_\nu(F, A)$.*

Hence the representations we have obtained are essentially unique.

Let K and K' be abstract rings. A mapping ϕ shall be called *a projection of the ideal theory* of K upon that of K' if ϕ is at the same time a projectivity of the set of right ideals of K upon the set of right ideals of K', the set of left ideals of K upon the set of left ideals of K', and the set of two-sided ideals of K upon those of K' and which also satisfies $(LR)^\phi = L^\phi R^\phi$ if L is any left ideal, and R any right ideal of K. (The statement that J is a left (right) ideal is not meant to preclude the possibility that J is a two-sided ideal).

If ϕ is a projection of the ideal theory of K upon that of K' it follows that ϕ maps right (left) annulets upon right (left) annulets, and in particular the right annihilator of a left ideal onto the right annihilator of its image. If I is a two-sided ideal, setting $L = R = I$ in condition above gives

$(I^2)^\phi = (I^\phi)^2$ so that a two-sided ideal which is not a zero ring is mapped upon a two-sided ideal which is not a zero ring.

If there exists a projection of the ideal theory of K upon the ideal theory of K', we shall say that the *rings K and K' have the same ideal theory.*

THEOREM 8. 2. *Assume the linear manifold (F, A) has rank at least three, and that the ring K possesses an identity element. Then the ring K and the ring of linear transformations $T(F, A)$ have the same ideal theory if and only if they are isomorphic. Moreover every projection of the ideal theory of K upon the ideal theory of $T(F, A)$ is induced by a unique ring isomorphism of K upon $T(F, A)$.*

Proof. Isomorphic rings have the same ideal theory. Now assume K and $T(F, A)$ have the same ideal theory. It follows from Theorem 7. 5 that $T(F, A)$ satisfies the condition (1), (2), (3) and (4) of that theorem. From our preceding remarks it follows that the ring K also satisfies the same conditions. Hence, Theorem 7. 5 then implies that there exists a linear manifold (G, B) so that K is isomorphic to the ring $T(G, B)$ of all linear transformations of B. It now follows from Theorem 2. 8 that there exists a projectivity of the system of subspaces of the linear manifold (F, A) upon the totality of left annulets of $T(F, A)$, and a projectivity of the system of subspaces of (G, B) upon the totality of left annulets of $T(G, B)$. Since $T(F, A)$ and $T(G, B)$ have the same ideal theory, there exists a projectivity of the totality of left annulets of $T(G, B)$ upon the totality of left annulets of $T(F, A)$. Thus, there exists a projectivity of (F, A) upon (G, B). Since $r(A) \geqq 3$ it follows from The Fundamental Theorem of Projective Geometry (Baer, [4], p. 44) that this latter projectivity is induced by a semi-linear transformation σ of (F, A) upon (G, B). If $\eta \in T(F, A)$, it is easily verified that the correspondence of η and $\sigma^{-1}\eta\sigma$ is an isomorphism of $T(F, A)$ upon $T(G, B)$ which is isomorphic to K. Now let the product of the projection of $T(F, A)$ upon K by the isomorphism of K upon $T(G, B)$ map the left annulet $L(S)$ of $T(F, A)$ upon the left annulet $L(U)$ of $T(G, B)$. Then the semi-linear transformation σ satisfies $S^\sigma = U$. But the equivalence of the relations:

$$\eta \in L(S), \qquad A\eta \leqq S, \qquad B(\sigma^{-1}\eta\sigma) \leqq U, \qquad \sigma^{-1}\eta\sigma \in L(U)$$

shows that the isomorphism we have constructed has the desired effect on the left annulets. In a similar way it has the desired effect on the right annulets. Since every ideal of $T(F, A)$ is a sum of annulets, (Remark 2 following Corollary 4. 5) it follows that the ring isomorphism induces the projection of the ideal theory of $T(F, A)$ upon the ideal theory of K.

Now let α, β be two isomorphisms which induce the same projection of the ideal theory of $T(F, A)$ upon that of K. Then $\alpha\beta^{-1}$ is an automorphism of $T(F, A)$ which, in particular, leaves invariant every left annulet of $T(F, A)$. Hence $\alpha\beta^{-1} = 1$ ([4], p. 187) and $\alpha = \beta$. This completes the proof of the theorem.

Remark 1. If the linear manifold (F, A) has rank less than three, the theorem is no longer valid. Let (F, A) and (G, B) have rank one. Then the rings $T(F, A)$ and $T(G, B)$ are essentially the same as the fields F and G respectively, and hence have the same ideal theory. But the fields F and G need not be isomorphic. For the case of rank two, we may let (F, A) and (G, B) be linear manifolds of rank two which are projectively equivalent, but such that there exists no semi-linear transformation of (F, A) upon (G, B). Examples of such manifolds are given in [4], pp. 50-51. By Theorem 2. 8 it follows that there exists a projectivity of the totality of left (right) annulets of the ring $T(F, A)$ upon the left (right) annulets of the ring $T(G, B)$. Since the ranks are finite, every ideal is an annulet. Using the fact that these rings are simple it is easily verified that these projectivities actually constitute a projection of the ideal theory of $T(F, A)$ upon that of $T(G, B)$. The rings $T(F, A)$ and $T(G, B)$ however cannot be isomorphic, since according to Theorem 8. 1 such an isomorphism is induced by a semi-linear transformation of (F, A) upon (G, B).

Since the correspondence of η and $\sigma^{-1}\eta\sigma$ of the theorem clearly maps linear transformations of finite rank, upon linear transformations of finite rank we also have the following:

COROLLARY 8. 3. *Assume the linear manifold (F, A) has rank at least three. Then every projection of the ideal theory of $T_0(F, A)$ upon the ideal theory of an abstract ring K is induced by a unique ring isomorphism.*[1]

THE UNIVERSITY OF ILLINOIS,
RUTGERS UNIVERSITY.

---

[1] This paper is a revised portion of a doctoral dissertation written at the University of Illinois under the direction of Professor Reinhold Baer. The author wishes to express his sincere appreciation to Professor Baer for his constant aid and encouragement.

BIBLIOGRAPHY.

[1] E. Artin and G. Whaples, "The theory of simple rings," *American Journal of Mathematics*, vol. 65 (1943), pp. 87-107.

[2] R. Baer, "Automorphism rings of primary Abelian operator groups," *Annals of Mathematics*, vol. 44 (1943), pp. 192-227.

[3] ———, "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.

[4] ———, *Linear algebra and projective geometry*, Academic Press, New York, 1952.

[5] G. Birkhoff, *Lattice theory*, American Mathematical Society Colloquium Publications, vol. 25 (1948).

[6] B. Brown and N. H. McCoy, "The maximal regular ideal of a ring," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 165-171.

[7] J. Dieudonné, "Sur le socle d'un anneau et les anneaux simples infinis," *Bulletin de la Société Mathématique de France*, vol. 70 (1942), pp. 46-75.

[8] N. Jacobson, *The theory of rings*, Mathematical Surveys, vol. 2, New York, 1943.

[9] ———, "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.

[10] ———, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

[11] ———, "On the theory of primitive rings," *Annals of Mathematics*, vol. 48 (1947), pp. 8-21.

[12] R. E. Johnson, "Equivalence rings," *Duke Mathematical Journal*, vol. 15 (1948), pp. 787-793.

[13] ——— and F. Kiokemeister, "The endomorphisms of the total operator domain of an infinite module," *Transactions of the American Mathematical Society*, vol. 62 (1947), pp. 404-430.

[14] I. Kaplansky, "Dual rings," *Annals of Mathematics*, vol. 49 (1948), pp. 689-701.

[15] ———, *The theory of rings*, University of Chicago (mimeographed), Winter 1950.

[16] J. von Neumann, "On regular rings," *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 707-713.

[17] ———, *Continuous geometry*, vol. 2, Princeton Lectures, 1936-37.

SYMMETRIC AND ANTI SYMMETRIC KRONECKER SQUARES AND INTERTWINING NUMBERS OF INDUCED REPRESENTATIONS OF FINITE GROUPS.*

By George W. Mackey.

Introduction. This paper is a continuation of part I of an earlier article [4]. There a number of results of Frobenius, Shoda and Artin were unified by deriving them as corollaries of a theorem on the structure of the Kronecker product of two induced representations. In the first half of the present paper we supplement the main theorem of [4] with parallel results on the symmetric and anti symmetric components of the Kronecker square of a single induced representation. These results of course have corollaries about the symmetric and anti symmetric intertwining numbers of pairs of adjoint induced representations. The special cases which deal with the self intertwining numbers of a permutation representation admit simple direct proofs and in the second half of the paper we use the results in these cases and the methods and results of [4] to extend the unification given in [4] so as to encompass certain results of Frame [1], [2] and Wigner [6]. Wigner's results are extended somewhat and, we think, made to seem less mysterious. A reader interested only in the applications in the second half may proceed directly from Section **1** to Theorem 2′ in Section **2**; going back however to read Corollary 2 to Theorem 2 (which is also a corollary to Theorem 2′) and the discussion following this corollary.

As with the results of [4] extensions are possible to Hilbert space representations of locally compact groups. We shall develop these extensions in detail elsewhere and confine ourselves here to a few indicatory remarks in a final paragraph.

We shall assume that the reader is familiar with [4] and shall use terminology and notation introduced therein without further explanation. We shall not in general distinguish between equivalent representations and shall use the term "restriction of a representation" in each of the following contexts: (a) To denote the representation of a subgroup obtained from a representation of a group by ignoring values of the group variable outside of the subgroup. (b) To denote the representation defined by an invariant sub-

* Received May 23, 1952.

space of another representation by ignoring what the linear transformations do outside of this subspace. We shall speak respectively of the restriction to the given subgroup and the restriction to the given subspace. In one case we change the group being represented and in the other the representation space.

1. Symmetry and anti symmetry of Kronecker squares and intertwining operators. Let $\mathcal{G}$ be a finite group and let $U, x \to U_x$, be an arbitrary representation of $\mathcal{G}$ by linear transformations in a vector space $\mathcal{H}(U)$ over a field $\mathcal{F}$ of odd characteristic. The space of the Kronecker square $U \otimes U$ of U is then the set of all linear transformations T from $\overline{\mathcal{H}(U)}$ to $\mathcal{H}(U)$ and $(U \otimes U)_x(T) = U_x T U^*_x$. Hence

$$((U \otimes U)_x(T))^* = (U_x T U^*_x)^* = U_x T^* U^*_x = (U \otimes U)_x(T^*).$$

Thus the involutory linear transformation $T \to T^*$ commutes with $(U \otimes U)_x$ for all x. It follows that the subspaces defined by the equations $T = T^*$ and $T = -T^*$ are invariant and define a two term direct sum decomposition of $U \otimes U$. Thus $U \otimes U = U \circledS U + U \circledA U$ where $U \circledS U$ is obtained by restricting $U \otimes U$ to the set of all $T \varepsilon \mathcal{H}(U \otimes U)$ with $T = T^*$ and $U \circledA U$ is obtained by restricting $U \otimes U$ to the set all $T \varepsilon \mathcal{H}(U \otimes U)$ with $T = -T^*$. We shall call $U \circledS U$ and $U \circledA U$ respectively the symmetric and anti symmetric Kronecker squares of U.

An intertwining operator T for $\bar{U}$ and U is a member of $\mathcal{H}(U \otimes U)$ which is carried into itself by all $(U \otimes U)_x$. It is easy to see that whenever T is an intertwining operator then T^* is also. Thus the space of intertwining operators for $\bar{U}$ and U splits as a direct sum of two subspaces and we have $\mathcal{J}(\bar{U}, U) = \mathcal{J}_S(\bar{U}, U) + \mathcal{J}_A(\bar{U}, U)$ where $\mathcal{J}_S(\bar{U}, U)$ is the dimension of the space of all intertwining operators T for which $T = T^*$ and $\mathcal{J}_A(\bar{U}, U)$ is the dimension of the space of all intertwining operators T for which $T = -T^*$. We call $\mathcal{J}_S(\bar{U}, U)$ and $\mathcal{J}_A(\bar{U}, U)$ the symmetric and anti symmetric intertwining numbers of $\bar{U}$ and U respectively. It is clear that $\mathcal{J}_S(\bar{U}, U) = \mathcal{J}(I, U \circledS U)$ and $\mathcal{J}_A(\bar{U}, U) = \mathcal{J}(I, U \circledA U)$ where I is the one dimensional identity representation of $\mathcal{G}$.

LEMMA 1. *If U and V are arbitrary representations of $\mathcal{G}$ then*

$$(U + V) \circledS (U + V) = U \circledS U + V \circledS V + U \otimes V$$

and

$$(U + V) \circledA (U + V) = U \circledA U + V \circledA V + U \otimes V$$

where of course $V \otimes U$ could replace its equivalent $U \otimes V$.

Proof. It is obvious that $(U+V)\otimes(U+V)$ has a natural decomposition as a direct sum $U\otimes U+V\otimes V+U\otimes V+V\otimes U$ where $T\to T^*$ leaves $\mathfrak{H}(U\otimes U)$ and $\mathfrak{H}(V\otimes V)$ invariant and maps $\mathfrak{H}(U\otimes V)$ and $\mathfrak{H}(V\otimes U)$ linearly onto one another. Let $T=T_1, T_2, T_3, T_4$ be an arbitrary element in $\mathfrak{H}(U\otimes U)\oplus\mathfrak{H}(V\otimes V)\oplus\mathfrak{H}(U\otimes V)\oplus\mathfrak{H}(V\otimes U)$. Then $T^*=T^*_1, T^*_2, T^*_4, T^*_3$ so that $T=T^*$ if and only if $T^*_1=T_1$, $T^*_2=T_2$, $T^*_4=T_3$ and $T^*_3=T_4$ and $T=-T^*$ if and only if $T^*_1=-T_1$, $T^*_2=-T_2$, $T^*_3=-T_4$ and $T^*_4=-T_3$. It follows that the space of $(U+V)\ \text{Ⓢ}\ (U+V)$ is the set of all $T_1\,T_2, T_3, T^*_3$ where $T_1\,\varepsilon\,\mathfrak{H}(U\ \text{Ⓢ}\ U)$, $T_2\,\varepsilon\,\mathfrak{H}(V\ \text{Ⓢ}\ V)$ and $T_3\,\varepsilon\,\mathfrak{H}(U\otimes V)$. Thus

$$(U+V)\ \text{Ⓢ}\ (U+V)=U\ \text{Ⓢ}\ U+V\ \text{Ⓢ}\ V+W$$

where W is a representation which is equivalent to $U\otimes V$ *by* way of the mapping $T_3, T^*_3\to T_3$. $(U+V)\ \text{Ⓐ}\ (U+V)$ may be treated in parallel fashion.

Corollary.

$$\mathfrak{I}_S(\overline{U+V}, U+V)=\mathfrak{I}_S(\bar{U},U)+\mathfrak{I}_S(\bar{V},V)+\mathfrak{I}(\bar{U},V)$$

and

$$\mathfrak{I}_A(\overline{U+V}, U+V)=\mathfrak{I}_A(\bar{U},U)+\mathfrak{I}_A(\bar{V},V)+\mathfrak{I}(\bar{U},V).$$

Making use of the notation $c(U)=\mathfrak{I}_S(\bar{U},U)-\mathfrak{I}_A(\bar{U},U)$ we have

Corollary. $c(U+V)=c(U)+c(V)$.

Note that if U is irreducible and $\mathfrak{F}$ is algebraically closed then $\mathfrak{I}(\bar{U},U)=1$ or 0 according as U is or is not equivalent to $\bar{U}$ and that hence $c(U)=1$, -1, or 0. This invariant of irreducible representations was introduced by Frobenius and Schur in [3] in order to classify representations over the complex field according to their "reality." They show that $c(U)=1$, -1, or 0 according as U may be realized by real matrices, has a real character function but may not be realized by real matrices or has a non real character function.

Lemma 2. *If $\mathfrak{F}$ is algebraically closed and U and V are irreducible representations of $\mathfrak{G}_1$ and $\mathfrak{G}_2$ then $c(U\times V)=c(U)c(V)$.*

Proof. Clearly $\overline{U\times V}=\bar{U}\times\bar{V}$. Thus $U\times V=\overline{U\times V}$ if and only if $U=\bar{U}$ and $V=\bar{V}$. Thus $c(U\times V)=0$ if and only if $c(U)=0$ or $c(V)=0$; that is if and only if $c(U)c(V)=0$. If $U=\bar{U}$ and $V=\bar{V}$ let T be an intertwining operator for $\bar{U}$ and U and let S be an intertwining

operator for $\bar{V}$ and V. Then $T \times S$ is an intertwining operator for $\overline{U \times V}$ and $U \times V$. Moreover $(T \times S)^* = T^* \times S^*$. The truth of the lemma is now evident.

COROLLARY. *If $\mathcal{F}$ is algebraically closed and U and V are direct sums of irreducible representations then $c(U \times V) = c(U)c(V)$.*

LEMMA 3. $c(\bar{U}) = c(U)$ *for all* U.

Proof.

$$c(\bar{U}) = \mathcal{I}_S(U, \bar{U}) - \mathcal{I}_A(U, \bar{U}) = \mathcal{I}(I, \bar{U} \circledS \bar{U}) - \mathcal{I}(I, \bar{U} \textcircled{A} \bar{U})$$
$$= \mathcal{I}(I, \overline{U \circledS U}) - \mathcal{I}(I, \overline{U \textcircled{A} U}) = \mathcal{I}(I, U \circledS U) - \mathcal{I}(I, U \textcircled{A} U) = c(U).$$

LEMMA 4.[1] *Let $\mathcal{F}$ be algebraically closed. Let G be a subgroup of the finite group $\mathfrak{G}$ and let L be an irreducible representation of $\mathfrak{G}$. Suppose that M, the restriction of L to G, is a direct sum of non equivalent irreducible representations M_j. Then $c(M_j) = c(L)$ for all j.*

Proof. It is clear that every intertwining operator T for $\bar{M}$ and M is uniquely a sum of intertwining operators T_j where T_j is zero on $\overline{\mathcal{H}(M_1)} \oplus \overline{\mathcal{H}(M_2)} \oplus \cdots \overline{\mathcal{H}(M_{j-1})} \oplus \overline{\mathcal{H}(M_{j+1})} \oplus \cdots \overline{\mathcal{H}(M_n)}$ and has $\mathcal{H}(M_j)$ for its range. This is so in particular for any intertwining operator which intertwines $\bar{L}$ and L. But such an operator if not zero is non singular. Hence no T_j is zero and we have either $T = T^*$ and hence $T_j = T^*_j$ for all j or we have $T = -T^*$ and hence $T_j = -T^*_j$ for all j.

2. Symmetric and anti symmetric squares of induced representations. Let G be a subgroup of the finite group $\mathfrak{G}$ and let L be a representation of G. It follows from Theorem 2 of [4] that $U^L \otimes U^L$ is a direct sum over the $G:G$ double cosets of certain other induced representations. In this section we shall show that $U^L \circledS U^L$ and $U^L \textcircled{A} U^L$ are also direct sums of induced representations and describe these representations quite explicitly. We assume of course that the underlying field is of characteristic different from two. To begin with let us recall that via the natural mapping of $\mathfrak{G}$ on the diagonal subgroup $\tilde{\mathfrak{G}}$ of $\mathfrak{G} \times \mathfrak{G}$ $U^L \otimes U^L$ is the restriction to $\tilde{\mathfrak{G}}$ of the representation $U^L \times U^L$ of $\mathfrak{G} \times \mathfrak{G}$ and that by Lemma 2 of [4] $U^L \times U^L$ is equivalent to $U^{L \times L}$. Thus we may identify the space of $U^L \otimes U^L$ with the set of all functions $A, x, y \to A_{x,y}$ from $\mathfrak{G} \times \mathfrak{G}$ to the set of all linear

[1] A special case of this lemma is proved by essentially the same argument in [6].

operators from $\overline{\mathcal{H}(L)}$ to $\mathcal{H}(L)$ such that for all $x, y \varepsilon \mathcal{G} \times \mathcal{G}$ and all $\xi, \eta \varepsilon G \times G$

$$(\dagger) \qquad A_{\xi x, \eta y} = L_\xi A_{x,y} L^*_\eta.$$

Moreover it is not difficult to verify that the adjoint operation in the space of $U^L \otimes U^L$ goes over into the operation which takes A into A' where $A'_{x,y} = A^*_{y,x}$. Finally $((U^L \otimes U^L)_s(A))_{x,y} = A_{xs,ys}$ for all x, y, and s in $\mathcal{G}$. For each $G \times G : \tilde{\mathcal{G}}$ double coset d in $\mathcal{G} \times \mathcal{G}$ let $\mathcal{H}_d$ be the set of all A in $\mathcal{H}(U^{L \times L})$ which vanish outside of d. It is clear that $\mathcal{H}(U^{L \times L})$ is a direct sum over the double cosets of the $\mathcal{H}_d$ and that each $\mathcal{H}_d$ is invariant under the transformations $(U^L \otimes U^L)_s$. This decomposition of $U^L \otimes U^L$ is that described in Theorem 2 of [4]. In order to deal with $U^L \circledS U^L$ and $U^L \circledA U^L$ we must alter the decomposition so that it is invariant under $A \to A'$ as well. We note that $(\mathcal{H}_d)' = \mathcal{H}_{d'}$ where d' is the double coset consisting of all x, y with $y, x \varepsilon d$. Let us call d' the transpose of d. Further let us denote $d \cup d'$ by $\bar{d}$ and let $\mathcal{B}$ denote the set of all subsets of $\mathcal{G} \times \mathcal{G}$ of the form $\bar{d}$. If $b \varepsilon \mathcal{B}$ let $\mathcal{H}_b = \mathcal{H}_d$ or $\mathcal{H}_d \oplus \mathcal{H}_{d'}$ according as $d = d'$ or not. Then $\mathcal{H}(U^L \otimes U^L)$ is a direct sum of the $\mathcal{H}_b$ for $b \varepsilon \mathcal{B}$ and each $\mathcal{H}_b$ is invariant under $A \to A'$ as well as under the operators $(U^L \otimes U^L)_s$. Let $\mathcal{H}^+_b$ be the set of all $A \varepsilon \mathcal{H}_b$ with $A = A'$ and let $\mathcal{H}^-_b$ be the set of all $A \varepsilon \mathcal{H}$ with $A = -A'$. Let V^{b+} and V^{b-} be the restrictions to $\mathcal{H}^+_b$ and $\mathcal{H}^-_b$ of $U^L \otimes U^L$. Then $U^L \circledS U^L = \sum_{b \varepsilon \mathcal{B}} V^{b+}$ and $U^L \circledA U^L = \sum_{b \varepsilon \mathcal{B}} V^{b-}$. These are the decompositions with which our theorem deals and all that remains is to identify the summands with specific induced representations of $\mathcal{G}$. There are three cases according as b contains two distinct double cosets, coincides with $(G \times G)\mathcal{G}$ or is a single double coset distinct from $(G \times G)\tilde{\mathcal{G}}$. The first is readily disposed of. The second and third require a more elaborate discussion.

Case I. *b contains two distinct double cosets.* Then $\mathcal{H}_b = \mathcal{H}_d \oplus \mathcal{H}_{d'}$. Clearly $\mathcal{H}^+_b$ consists of all $A + A'$ where $A \varepsilon \mathcal{H}_d$ and $\mathcal{H}^-_b$ consists of all $A - A'$ where $A \varepsilon \mathcal{H}_d$. Consider the linear mappings $A \to A + A'$ and $A \to A - A'$. These map $\mathcal{H}_d$ in a one to one linear manner onto $\mathcal{H}^+_b$ and $\mathcal{H}^-_b$ respectively and since $A \to A'$ commutes with all $(U^L \otimes U^L)_s$ they set up equivalences between $U^L \otimes U^L$ restricted to $\mathcal{H}_d$ and V^{b+} and V^{b-} respectively. Thus the four representations V^{b+}, V^{b-}, $U^L \otimes U^L$ restricted to $\mathcal{H}_d$ and $U^L \otimes U^L$ restricted to $\mathcal{H}_{d'}$ are all equivalent. But the latter two representations have already been identified in Theorem 2 of [4]. Indeed choose any x, y in b. Let xLx^{-1} and yLy^{-1} denote the representations $\xi \to L_{x\xi x^{-1}}$, $\eta \to L_{y\eta y^{-1}}$ of $x^{-1}Gx \cap y^{-1}Gy$. Let $M = (xLx^{-1}) \otimes (yLy^{-1})$. Then it follows

from the preceding discussion and Theorem 2 of [4] that V^{b+} and V^{b-} are both equivalent to the representation U^M of $\mathfrak{G}$ induced by the representation M of $x^{-1}Gx \cap y^{-1}Gy$.

Cases II *and* III. *b consists of a single double coset d with $d = d'$.* Let x, y be an arbitrary element of d. It follows from (†) that each A in $\mathfrak{H}_b = \mathfrak{H}_d$ is determined throughout b by its values on the left coset $(x, y)\tilde{\mathfrak{G}}$. Given A let us define a function B^A on $\mathfrak{G}$ as follows: $B^A{}_t = A_{xt,yt}$ for all t in $\mathfrak{G}$. Then (x, y being held fixed) B^A is determined by A and in turn determines A via the equation:

$$(\dagger\dagger) \qquad A_{\xi xt,\eta yt} = L_\xi B^A{}_t L^*{}_\eta$$

valid for all $\xi,\eta \in G \times G$ and all $t \in \mathfrak{G}$. A straightforward calculation now shows that an arbitrary function B from $\mathfrak{G}$ to $\mathfrak{H}(L \otimes L)$ is of the form B^A for some A in $\mathfrak{H}_b$ if and only if

$$(\dagger\dagger\dagger) \qquad B_{\zeta t} = L_{x\zeta x^{-1}}(B_t) \; L^*{}_{y\zeta y^{-1}}$$

for all $\zeta \in x^{-1}Gx \cap y^{-1}Gy$ and that moreover if $A_1 = (U^L \otimes U^L)_s A$ then $B^{A_1}{}_t = B^A{}_{ts}$. We now investigate the effect on B^A of replacing A by A'. That is given B we find A so that $B = B^A$ and then express $B^{A'}$ in terms of B. Since $A'_{x,y} = A^*{}_{y,x}$ we have $B'_t =$ (by definition) $B^{A'}{}_t = A'_{xt,yt}$ $= A^*{}_{yt,xt}$. If we can find $\xi, \eta \in G \times G$ and a mapping s of $\mathfrak{G}$ into $\mathfrak{G}$ such that $\xi xs(t) = yt$ and $\eta ys(t) = xt$ for all $t \in \mathfrak{G}$ then

$$\begin{aligned} B^{A'}{}_t = A^*{}_{\xi xs(t),\eta ys(t)} &= (L_\xi(A_{xs(t),ys(t)})L^*{}_\eta)^* \\ &= (L_\xi B^A{}_{s(t)} L^*{}_\eta)^* = L_\eta (B^A{}_{s(t)})^* L^*{}_\xi. \end{aligned}$$

Thus $A \to A'$ will go over into $B \to B'$ where $B'_t = L_\eta B^*{}_{s(t)} L^*{}_\xi$. Now the equations $\xi xs(t) = yt$ and $\eta ys(t) = xt$ are equivalent to $s(t) = x^{-1}\xi^{-1}yt$ $= y^{-1}\eta^{-1}xt$. Thus we must choose ξ and η so that $x^{-1}\xi^{-1}y = y^{-1}\eta^{-1}x$ and then set $s(t) = zt$ where $z = x^{-1}\xi^{-1}y = y^{-1}\eta^{-1}x$. But $x^{-1}\xi^{-1}y = y^{-1}\eta^{-1}x$ if and only if $\eta yx^{-1}\xi^{-1} = xy^{-1}$; that is if and only if xy^{-1} and its inverse are in the same $G : G$ double coset. Now as was pointed out in [4] the $G : G$ double cosets in $\mathfrak{G}$ and the $G \times G : \tilde{\mathfrak{G}}$ double cosets in $\mathfrak{G} \times \mathfrak{G}$ are in one-to-one correspondence in such a manner that $G \times G(x, y)\tilde{\mathfrak{G}}$ corresponds to $Gxy^{-1}G$ and it is clear that xy^{-1} lies in a self inverse $G : G$ double coset if and only if x, y and y, x lie in the same $G \times G : \tilde{\mathfrak{G}}$ double coset. In short ξ and η may be found if and only if b consists of a single double coset and this is the case under discussion. Thus $x^{-1}Gy \cap y^{-1}Gx$ is not empty and if we let z be any one of its elements we have $B'_t = L_\eta B^*{}_{zt} L^*{}_\xi$ where

$\eta = xz^{-1}y^{-1}$ and $\xi = yz^{-1}x^{-1}$. It is convenient to summarize the argument up to this point in a lemma.

LEMMA a. *If d is a $G \times G:\tilde{\mathfrak{G}}$ double coset such that $d = d'$ and x, y is any element of d then the representation of $\mathfrak{G}$ obtained by restricting $U^L \otimes U^L$ to $\mathfrak{H}_d$ is equivalent to the representation U^M where M is the representation $(xLx^{-1}) \otimes (yLy^{-1})$ of $x^{-1}Gx \cap y^{-1}Gy$. Moreover $x^{-1}Gy \cap y^{-1}Gx$ is not empty and if z is any one of its elements then the involution $B \to B'$ whose 1 and -1 spaces are the spaces of V^{b+} and V^{b-} respectively takes the following form:*

$$B'_t = L_{xz^{-1}y^{-1}}(B^*_{zt})L^*_{yz^{-1}x^{-1}}.$$

We also have

LEMMA b. *Let d, x, y and z be as in Lemma a and let $G_0 = x^{-1}Gx \cap y^{-1}Gy$. Then $z^2 \varepsilon G_0$ and $zG_0z^{-1} = G_0$ so that z and G_0 generate a subgroup G_1 of $\mathfrak{G}$ which contains G_0 as a normal subgroup of index two.*

Proof. Since

$$z \,\varepsilon\, x^{-1}Gy \cap y^{-1}Gx, \quad z^2 \,\varepsilon\, x^{-1}Gyy^{-1}Gx = x^{-1}Gx \text{ and } z^2 \,\varepsilon\, y^{-1}Gxx^{-1}Gy = y^{-1}Gy.$$

Therefore $z^2 \,\varepsilon\, G_0$. Moreover

$$z(x^{-1}Gx)z^{-1} = y^{-1}\xi xx^{-1}Gxx^{-1}\xi^{-1}y = y^{-1}Gy$$

and

$$z(y^{-1}Gy)z^{-1} = x^{-1}\eta yy^{-1}Gyy^{-1}\eta^{-1}x = x^{-1}Gx$$

where $\xi \,\varepsilon\, G$ and $\eta \,\varepsilon\, G$. Thus the inner automorphism defined by z interchanges the two groups of which G_0 is the intersection and hence leaves G_0 invariant.

LEMMA c. *Let the terminology be as in the two preceding lemmas. Then $G_0 = G_1$ if and only if $d = (G \times G)\tilde{\mathfrak{G}}$.*

Proof. Let $d = (G \times G)\tilde{\mathfrak{G}}$. Then $x = \xi a$ and $y = \eta a$ where $a \,\varepsilon\, \mathfrak{G}$ and $\xi, \eta \,\varepsilon\, G \times G$. Hence

$$x^{-1}Gy \cap y^{-1}Gx = a^{-1}Ga \cap a^{-1}Ga = a^{-1}Ga$$

and $G_0 = a^{-1}Ga \cap a^{-1}Ga = a^{-1}Ga$. It follows that $z \,\varepsilon\, G_0$ and $G_0 = G_1$. Conversely if $G_0 = G_1$ so that $z \,\varepsilon\, G_0$ then there exists $\xi \,\varepsilon\, G$ so that $x^{-1}\xi y \,\varepsilon\, x^{-1}Gx \cap y^{-1}Gy$. Hence $y \,\varepsilon\, Gx$. Hence $x, y \,\varepsilon\, (G \times G)\tilde{\mathfrak{G}}$. Therefore $d = (G \times G)\tilde{\mathfrak{G}}$.

We now discuss cases II and III separately.

Case II. $d = (G \times G)\tilde{\mathfrak{G}}$. Here we may choose $x = y = z = e$ where e

is the identity of $\mathfrak{G}$ and the expression for B'_t in Lemma a simplifies to $B'_t = B^*_t$. Thus $B = B'$ if and only if $B_t = B^*_t$ for all t and $B' = -B$ if and only if $B_t = -B^*_t$ for all t. It should now be evident that in this case V^{b+} and V^{b-} are equivalent to $U^{L \circledS L}$ and $U^{L \circledA L}$ respectively.

Case III. $d = d'$ but $d \neq (G \times G)\tilde{\mathfrak{G}}$. In this case $z \notin G_0$ and G_0 is a proper subgroup of G_1. We know that $V^{b+} + V^{b-} = U^M$ where M is the representation $(xLx^{-1}) \otimes (yLy^{-1})$ of G_0. We shall identify V^{b+} and V^{b-} by proving them equivalent to the induced representations U^{M^+} and U^{M^-} where M^+ and M^- are certain extensions of M to the group G_1. To this end let T be the non singular linear operator in $\mathcal{H}(L \otimes L) = \mathcal{H}(M)$ which takes S into $L_{xz^{-1}y^{-1}}(S^*)\, L^*_{yz^{-1}x^{-1}}$. Then

$$T^2(S) = L_{xz^{-1}y^{-1}} L_{yz^{-1}x^{-1}}(S) L^*_{xz^{-1}y^{-1}} L^*_{yz^{-1}x^{-1}} = L_{xz^{-2}x^{-1}}(S) L_{yz^{-2}y^{-1}} = M_{z^{-2}}(S).$$

Thus $T^2 = M_{z^{-2}}$ and this suggests that we might be able to extend M to G_1 by defining M_z to be T^{-1} or $-T^{-1}$. Of course an extension of M to G_1 is uniquely determined by its value Q at z and if such an extension exists we surely have $Q^2 = M_{z^2}$ and $M_{z\zeta z^{-1}} Q M_\zeta Q^{-1}$ for all $\zeta \in G_0$. Conversely it is easy to see that given any Q satisfying these two conditions there exists a unique extension of M to G_1 such that $M_z = Q$. Thus if we can verify that $M_{z\zeta z^{-1}} = T^{-1} M_\zeta T$ for all $\zeta \in G_0$ we will be assured that there exist unique extensions M^+ and M^- of M to G_1 such that $M^+_z = T^{-1}$ and $M^-_z = -T^{-1}$. But

$$\begin{aligned}(T^{-1} M_\zeta T)(S) &= T^{-1} M_\zeta (L_{xz^{-1}y^{-1}}(S^*) L^*_{yz^{-1}x^{-1}}) = T^{-1}(L_{x\zeta x^{-1}} L_{xz^{-1}y^{-1}}(S^*) \\ &\quad L^*_{yz^{-1}x^{-1}} L^*_{y\zeta y^{-1}}) = T^{-1}(L_{x\zeta z^{-1}y^{-1}} S^* L^*_{y\zeta z^{-1}x^{-1}}) = (L_{yzx^{-1}} L_{x\zeta z^{-1}y^{-1}}(S^*) \\ &\quad L^*_{y\zeta z^{-1}x^{-1}} L^*_{xyz^{-1}})^* = L_{xz\zeta z^{-1}x^{-1}}(S) L^*_{yz\zeta z^{-1}y^{-1}} = M_{z\zeta z^{-1}}(S).\end{aligned}$$

Thus T has the required properties and may be extended to M^+ and M^- as indicated. Finally note that $B'_t = T(B_{zt})$ for all t. Thus $B'_t \equiv B_t$ if and only if $B_{zt} \equiv T^{-1}(B_t)$; that is $B_{zt} \equiv M^+_z(B_t)$. Similarly $B' \equiv -B_t$ if and only if $B_{zt} \equiv M^-_z(B_t)$. Since we know already that $B_{\zeta t} = M_\zeta(B_t)$ for all $t \in \mathfrak{G}$ and all $\zeta \in G_0$ we see finally that $\mathcal{H}(V^{b+})$ is the set of all functions B from $\mathfrak{G}$ to $\mathcal{H}(M^+)$ such that $B_{\theta t} = M^+_\theta(B_t)$ for all $\theta \in G_1$ and all $t \in \mathfrak{G}$ and that $\mathcal{H}(V^{b-})$ is similarly related to M^-. In other words we see that V^{b+} and V^{b-} are equivalent respectively to the induced representations U^{M^+} and U^{M^-}. We have now completed the proof of our main theorem which we may state as follows.

THEOREM 1. *Let L be a representation of the subgroup G of the finite group $\mathfrak{G}$ and let the field of the vector space $\mathcal{H}(L)$ be of characteristic*

different from two. Let $\mathscr{B}_1$ be the set of all self inverse $G:G$ double cosets in $\mathscr{G}$ except G itself. Let $\mathscr{B}_2$ be the set of all sets of the form $d \cup d'$ where d is a non self inverse $G:G$ double coset and d' is the inverse of d. For each $b \in \mathscr{B}_2$ choose x and y so that $xy^{-1} \in b$ and let M be the representation $xLx^{-1} \otimes yLy^{-1}$ of $G_0 = x^{-1}Gx \cap y^{-1}Gy$. Then the induced representation U^M of $\mathscr{G}$ is independent of the choice of x and y and may be denoted by V^b. For each $b \in \mathscr{B}_1$ choose x and y so that $xy^{-1} \in b$. Then $x^{-1}Gy \cap y^{-1}Gx$ is not empty. Let z be any one of its members. Let G_0 be as defined above and let G_1 be the subgroup generated by G_0 and z. Then G_0 is a normal subgroup of G_1 of index two. Let M be the representation of G_0 defined as described above and let T be the linear transformation in $\mathscr{H}(M)$ which takes S into $L_{xz^{-1}y^{-1}}(S^*)L^*_{yz^{-1}x^{-1}}$. Then there exists unique extensions M^+ and M^- of M to G_1 such that $M^+_z = T^{-1}$ and $M^-_z = -T^{-1}$. The induced representations U^{M^+} and U^{M^-} are independent of x, y and z and may be denoted by V^{b+} and V^{b-} respectively. Finally we have

$$U^L \circledS U^L = U^{L \circledS L} + \sum_{b \in \mathscr{B}_1} V^{b+} + \sum_{b \in \mathscr{B}_2} V^b,$$

$$U^L \circledA U^L = U^{L \circledA L} + \sum_{b \in \mathscr{B}_1} V^{b-} + \sum_{b \in \mathscr{B}_2} V^b.$$

COROLLARY 1. *The symmetric (resp. anti symmetric) Kronecker square of an induced representation is equivalent to a direct sum of induced representations.*

When L is one dimensional (that is, when U^L is monomial) so that L may be regarded as an $\mathscr{F}$ valued function the descriptions of M^+ and M^- may be somewhat simplified. M is then also an $\mathscr{F}$ valued function and it may be extended to G_1 so as to be a homomorphism of G_1 into $\mathscr{F}$ in exactly two ways; namely by defining its value at z to be $L(xz^2x^{-1})$ or to be $-L(xz^2x^{-1})$. These two extensions are the M^+ and M^- of the theorem. We note correspondingly the corollary.

COROLLARY 2. *The symmetric (respectively anti symmetric) Kronecker square of a monomial representation is equivalent to a direct sum of monomial representations.*

Because of the relationship between intertwining operators and Kronecker products it is easy to deduce from Theorem 1 formulae for computing $\mathscr{I}_S(\overline{U^L}, U^L)$ and $\mathscr{I}_A(\overline{U^L}, U^L)$. The deduction is straight-forward and we shall content ourselves with a statement of the result. Moreover in order to avoid excessively complicated statements we shall confine ourselves to the case in which L is one dimensional and hence a character of G.

THEOREM 2. *Let L, G, $\mathfrak{G}$, $\mathfrak{F}$, $\mathfrak{B}_1$, $\mathfrak{B}_2$, be as in Theorem 1 except for the additional assumption that L is one dimensional. For each $b \in \mathfrak{B}_2$ choose x and y so that $xy^{-1} \in b$ and let M be the character $(xLx^{-1})(yLy^{-1})$. Then whether or not $M(\zeta) \equiv 1$ depends only upon b and we may write $j(b) = 1$ if $M(\zeta) \equiv 1$ and $j(b) = 0$ if $M(\zeta) \not\equiv 1$. For each $b \in \mathfrak{B}_1$ choose x and y so that $xy^{-1} \in b$. Then there exists $z \in x^{-1}Gy \cap y^{-1}Gx$ and if G_1 is the subgroup generated by $G_0 = x^{-1}Gx \cap y^{-1}Gy$ and z then G_0 is a normal subgroup of G_1 of index two. If M is defined as above then M may be extended so as to be a character of G_1 in exactly two ways as follows: $M^+(z) = L(xz^2x^{-1})$, $M^-(z) = -L(xz^2x^{-1})$. Whether or not $M(\zeta) \equiv 1$ depends only upon b. Moreover if $M(\zeta) \equiv 1$ so that $L(xz^2x^{-1}) = \pm 1$ then whether the plus or minus sign occurs depends only upon b. If $M(\zeta) \not\equiv 1$ we set $j(b) = 0$. If $M(\zeta) \equiv 1$ we set $j(b) = L(xz^2x^{-1})$. Then*

$$\mathfrak{I}_S(\overline{U^L}, U^L) = \mathfrak{I}_S(\bar{L}, L) + \sum_{b \in \mathfrak{B}_1} \tfrac{1}{2}(j(b)(1 + j(b))) + \sum_{b \in \mathfrak{B}_2} j(b),$$

$$\mathfrak{I}_A(\overline{U^L}, U^L) = \mathfrak{I}_A(\bar{L}, L) + \sum_{b \in \mathfrak{B}_1} \tfrac{1}{2}(j(b)(-1 + j(b))) + \sum_{b \in \mathfrak{B}_2} j(b).$$

Subtracting the two equations of the conclusion of the theorem we find the corollary:

COROLLARY 1. $c(U^L) = c(L) + \sum_{b \in \mathfrak{B}_1} j(b)$.

If in particular L is the identity so that U^L is a permutation representation then $j(b) = 1$ for all $b \in \mathfrak{B}_1$ and we have

COROLLARY 2. *If L is the identity representation of the subgroup G of $\mathfrak{G}$ then $c(U^L)$ is equal to the number of self inverse $G\colon G$ double cosets in $\mathfrak{G}$.*

In the special case in which U^L is completely reducible we know from the second corollary of Lemma 1 that $c(U^L)$ is the sum of the multiplicities of those irreducible components V of U^L for which $c(V) = 1$ minus the sum of the multiplicities of those irreducible components V of U^L for which $c(V) = -1$. Thus Corollary 2 includes the first main theorem of Frame's paper [1]. It should be noted that the special case of Theorem 2 in which L is the identity representation can be proved directly quite easily without appeal to the somewhat complicated general Theorem 1. Since the applications that we give in the present paper involve only this special case it seems worthwhile to give this proof explicitly.

THEOREM 2′. *Let G be a subgroup of the finite group $\mathfrak{G}$. Let I be the identity representation of G with respect to a field $\mathfrak{F}$ whose characteristic is not equal to two. Let n_1 denote the number of self inverse $G\colon G$ double cosets*

in $\mathfrak{G}$ and let n_2 denote the number of non self inverse $G:G$ double cosets in $\mathfrak{G}$. Then $\mathfrak{I}_S(U^I, U^I) = n_1 + \frac{1}{2}(n_2)$ and $\mathfrak{I}_A(U^I, U^I) = \frac{1}{2}(n_2)$.

Proof. (direct) By the argument at the beginning of section two, the general linear transformation from $\mathfrak{H}(\overline{U^I})$ into $\mathfrak{H}(U^I)$ will be defined by an $\mathfrak{F}$ valued function A on $\mathfrak{G} \times \mathfrak{G}$ such that $A_{\xi x, \eta y} = A_{x,y}$ for all $\xi, \eta \varepsilon G \times G$ and all $x, y \varepsilon \mathfrak{G} \times \mathfrak{G}$. This operator will be an intertwining operator if and only if $A_{xs,ys} = A_{x,y}$ for all x, y, and s in $\mathfrak{G}$. In short the intertwining operators correspond in a one-to-one linear fashion to the $\mathfrak{F}$ valued functions on $\mathfrak{G} \times \mathfrak{G}$ which are constant on the $G \times G : \tilde{\mathfrak{G}}$ double cosets—or simply to the $\mathfrak{F}$ valued functions on the $(G \times G) : \tilde{\mathfrak{G}}$ double cosets. In this correspondence the adjoint of the operation defined by $d \to A(d)$ is simply $d \to A(d')$ where d' is the set of all x, y with $y, x \varepsilon d$. Since $A(d) = -A(d')$ is impossible when $d = d'$ unless $A(d') = 0$ the theorem is now an obvious consequence of the correspondence between $G:G$ double cosets in $\mathfrak{G}$ and $G \times G : \tilde{\mathfrak{G}}$ double cosets in $\mathfrak{G} \times \mathfrak{G}$.

3. Simply reducible groups. Wigner in [6] has defined a finite group $\mathfrak{G}$ to be *simply reducible* if it has the following two properties: (a) Every representation L is equivalent to its adjoint $\bar{L}$. (b) Every Kronecker product $L \otimes M$ of irreducible representations is a direct sum of irreducible representations each of which occurs with multiplicity one. The main result of [6] is the following curious characterization of simply reducible groups. $\mathfrak{G}$ is simply reducible if and only if $\sum_{x \varepsilon G} \zeta(x)^3 = \sum_{x \varepsilon G} v(x)^2$ where for each x in $\mathfrak{G}$, $\zeta(x)$ is the number of square roots of x and $v(x)$ is the number of elements which commute with x. It is also shown in [6] that for arbitrary groups the left hand side of the above equation is less than or equal to the right hand side. In this section we shall establish a somewhat different characterization of simply reducible groups which is like Wigner's in that it is expressed in terms independent of representation theory but has the advantage that it is significant for infinite groups as well. In the next section we shall discuss the connection between the two characterizations and show how to derive one from the other. The chief tools in our approach are Theorem 2′ and the following easy consequence of Lemma 1. We shall assume from now on that $\mathfrak{F}$ is algebraically closed and of characteristic which is not equal to two and does not divide the order of $\mathfrak{G}$.

LEMMA 5. *Let $M = M_1 + M_2 + \cdots + M_n$ where the M_j are irreducible representations of the finite group $\mathfrak{G}$. Then $\mathfrak{I}_A(\bar{M}, M) = 0$ if and only if for each j either*

(a) $c(M_j) = 1$ and *M_j is not equivalent to M_k for any $k \neq j$*

or

(b) $c(M_j) = 0$ *and $\bar{M}_j$ is not a component of M.*

Proof. For each j let $M^0{}_j = M_1 + M_2 + \cdots M_{j-1} + M_{j+1} + \cdots + M_n$. By Lemma 1, $\mathfrak{I}_A(\bar{M}, M) = \mathfrak{I}_A(\bar{M}_j, M_j) + \mathfrak{I}_A(\bar{M}^0{}_j, M^0{}_j) + \mathfrak{I}(\bar{M}_j, M^0{}_j)$. Hence $\mathfrak{I}_A(\bar{M}, M) = 0$ if and only if all j we have $\mathfrak{I}_A(\bar{M}_j, M_j) = 0$ and for all k with M_k a component of $M^0{}_j$ we have $\mathfrak{I}(\bar{M}_j, M_k) = 0$. Now the first part of this last statement is equivalent to the statement that $c(M_j) = 1$ or 0. Moreover if $c(M_j) = 1$ then $M_j = \bar{M}_j$ so the second part is equivalent to the statement that M_j does not occur in $M^0{}_j$; that is that M_j occurs with multiplicity one. If $c(M_j) = 0$ the second part is equivalent to the statement that $\bar{M}_j$ does not occur at all. Thus the lemma is proved.

If $M = \bar{M}$ then alternative (b) is impossible and we have the

COROLLARY. *If $M = \bar{M}$ then $\mathfrak{I}_A(\bar{M}, M) = 0$ if and only if M is a direct sum of distinct irreducible components M_j with $c(M_j) = 1$.*

THEOREM 3. *If I is the identity representation of the subgroup G of the finite group $\mathfrak{G}$ then the following statements are equivalent:*

(a) $\mathfrak{I}_A(\overline{U^I}, U^I) = 0$.

(b) *Every $G : G$ double coset is self inverse.*

(c) *The irreducible components M_j of U^I occur with multiplicity one and for all j, $c(M_j) = 1$.*

Proof. The equivalence of (a) and (c) follows from the corollary to Lemma 5. The equivalence of (a) and (b) follows from Theorem 2′.

We obtain our characterization of simply reducible groups by applying Theorem 3 with a suitably chosen G and $\mathfrak{G}$. Let $\mathfrak{G}$ be an arbitrary finite group and let $\tilde{\mathfrak{G}}_3$ denote the diagonal subgroup of the triple Cartesian product $\mathfrak{G}_3 = \mathfrak{G} \times \mathfrak{G} \times \mathfrak{G}$ of $\mathfrak{G}$ with itself; that is let $\tilde{\mathfrak{G}}_3$ be the set of all x, y, z with $x = y = z$. Let I be the identity representation of $\tilde{\mathfrak{G}}_3$ and let U^I be the representation of $\mathfrak{G}_3$ induced by I. The general irreducible representation of $\mathfrak{G}_3$ is $L \times M \times N$ where L, M, and N are arbitrary irreducible representations of $\mathfrak{G}$. Moreover by the Frobenius reciprocity theorem (which may be obtained as a corollary of Theorem 2 of [4]) the multiplicity of occurrence of $L \times M \times N$ in U^I is equal to the multiplicity of occurrence of I in $L \otimes M \otimes N$; that is to the multiplicity of occurrence of $\bar{L}$ in $M \otimes N$.

Thus a necessary and sufficient condition that all irreducible components of U^I occur with multiplicity one is that for all irreducible representations M and N of $\mathfrak{G}$ all irreducible components of $M \otimes N$ occur with multiplicity one. In short $\mathfrak{G}$ satisfies (b) of the definition of simple reducibility if and only if U^I satisfies the first part of (c) under Theorem 3. On the other hand since $c(L \times M \times N) = c(L)c(M)c(N)$ a necessary and sufficient condition that the second part of (c) should be satisfied is that $c(L)c(M)c(N) = 1$ whenever L appears as an irreducible component of $M \otimes N$. But applying Lemma 4 to the restriction of $M \times N$ to the diagonal in $\mathfrak{G} \times \mathfrak{G}$ we see that this is the case whenever $M \otimes N$ is free of multiplicities and $c(M)c(N) \neq 0$; that is for all M and N if $\mathfrak{G}$ is simply reducible. Note next that if $c(L)c(M)c(N) = 1$ whenever L appears in $M \otimes N$ then $c(M) \neq 0$ for any M. Thus the second part of (c) under Theorem 3 implies (a) in the definition of simple reducibility. We have now proved that $\mathfrak{G}$ is simply reducible if and only if U^I satisfies (c) under Theorem 3. Applying Theorem 3 we conclude at once the truth of

THEOREM 4. *Let $\mathfrak{G}$ be a finite group and let $\tilde{\mathfrak{G}}_3$ be the diagonal subgroup in $\mathfrak{G} \times \mathfrak{G} \times \mathfrak{G}$. Then $\mathfrak{G}$ is simply reducible if and only if every $\tilde{\mathfrak{G}}_3 : \tilde{\mathfrak{G}}_3$ double coset in $\mathfrak{G} \times \mathfrak{G} \times \mathfrak{G}$ is self inverse.*

4. Wigner's condition. We shall relate our characterization of simply reducible groups to Wigner's by interpreting the two sides of his equation. This will lead us to a generalization of Wigner's inequality and to further theorems like Theorem 4.

LEMMA 6. *Let $\mathfrak{G}$ act as a group of permutations on a finite set S. Let $\mathfrak{G}$ contain h elements and let T be a permutation of S which commutes with the permutations defined by the members of $\mathfrak{G}$. For each $y \varepsilon \mathfrak{G}$ let $p(y)$ be the number of elements s of S such that $y(s) = T(s)$. Then $(1/h) \sum_{y \varepsilon G} p(y)$ is equal to the number of orbits of S under $\mathfrak{G}$ which are invariant under T.*

Proof. For each $s, y \varepsilon S \times \mathfrak{G}$ let $k(s,y) = 0$ or 1 according as $y(s) \neq T(s)$ or $y(s) = T(s)$. For each $s \varepsilon S$ let $q(s)$ be the number of elements y in $\mathfrak{G}$ such that $y(s) = T(s)$. Then

$$\sum_{y \varepsilon G} p(y) = \sum_{y \varepsilon G} \sum_{s \varepsilon S} k(s,y) = \sum_{s \varepsilon S} \sum_{y \varepsilon G} k(s,y) = \sum_{s \varepsilon S} q(s).$$

But $q(s)$ is zero if $T(s)$ is in the orbit containing s and h/n_s where n_s is the number of elements contained in orbit containing s if $T(s)$ is not in the orbit

containing s. Hence $\sum_{y \varepsilon G} p(y) = \sum'_{s \varepsilon S} (h/n_s)$ where the $'$ indicates that all s in non T invariant orbits have been omitted. Now for each T invariant orbit the term h/n_s occurs n_s times. Hence $\sum_{y \varepsilon G} p(y)$ is equal to h times the number of T invariant orbits as was to be proved.

THEOREM 5. *Let $\mathfrak{G}$ be a finite group of order h and let $\mathfrak{G}_{n+1}$ be the direct product of $\mathfrak{G}$ with itself $n+1$ times where $n \geqq 1$. Let $\tilde{\mathfrak{G}}_{n+1}$ be the diagonal subgroup of $\mathfrak{G}_{n+1}$. For each $x \varepsilon \mathfrak{G}$ let $v(x)$ be the number of elements of $\mathfrak{G}$ which commute with x. Then the following three numbers are equal:*

(a) $(1/h) \sum_{x \varepsilon G} v(x)^n$.

(b) *The number of $\tilde{\mathfrak{G}}_{n+1} : \tilde{\mathfrak{G}}_{n+1}$ double cosets in $\mathfrak{G}_{n+1}$.*

(c) *The number of orbits in $\mathfrak{G}_n$ under the group of inner automorphisms defined by members of the diagonal of $\mathfrak{G}_n$.*

Proof. The equality of (b) and (c) is an obvious consequence of their definitions and that of (a) and (c) follows at once from Lemma 6. We need only let $S = \mathfrak{G}_n$, $y(x_1, x_2, \cdots, x_n) = y^{-1}x_1y,\ y^{-1}x_2y, \cdots, y^{-1}x_ny$ and $T(x_1, x_2, \cdots, x_n) = x_1, x_2, \cdots, x_n$. $p(y)$ is clearly $v(y)^n$.

THEOREM 6. *Let $\mathfrak{G}$, $\mathfrak{G}_{n+1}$ and $\tilde{\mathfrak{G}}_{n+1}$ be as in Theorem 5. For each $x \varepsilon \mathfrak{G}$ let $\zeta(x)$ be the number of square roots of x. Then the following three numbers are all equal:*

(a) $(1/h) \sum_{x \varepsilon G} \zeta(x)^{n+1}$.

(b) *The number of self inverse $\tilde{\mathfrak{G}}_{n+1} : \tilde{\mathfrak{G}}_{n+1}$ double cosets in $\mathfrak{G}_{n+1}$.*

(c) *The number of self inverse orbits in $\mathfrak{G}_n$ under the group of inner automorphisms defined by members of the diagonal of $\mathfrak{G}_n$.*

Proof. The equality of (b) and (c) is an obvious consequence of the definitions. Applying Lemma 6 as in the proof of Theorem 5 but now taking $T(x_1, x_2, \cdots, x_n) = x_1^{-1}, x_2^{-1}, \cdots, x_n^{-1}$ we find that (c) is equal to $\sum_{y \varepsilon G} \zeta_1(y)^n$ where for each y in $\mathfrak{G}$, $\zeta_1(y)$ is the number of x in $\mathfrak{G}$ such that $y^{-1}xy = x^{-1}$. But $y^{-1}xy = x^{-1}$ if and only if $xyx = y$; that is if and only if $(xy)^2 = y^2$. Hence $\zeta_1(y) = \zeta(y^2)$. Hence $\sum_{y \varepsilon G} \zeta_1(y)^n = \sum_{y \varepsilon G} \zeta(y^2)^n = \sum_{z \varepsilon G} (\zeta(z))^n$ (number of y with $y^2 = z$) $= \sum_{z \varepsilon G} (\zeta(z))^{n+1}$.

As an immediate consequence of Theorems 4 and 5 we derive

THEOREM 7. *Let $\mathfrak{G}$ be any finite group and let v, ζ, $\mathfrak{G}_n$, $\tilde{\mathfrak{G}}_n$ be defined as in Theorems 5 and 6. Then for all $n = 1, 2, \cdots$ we have*

$$\sum_{x \varepsilon G} \zeta(x)^{n+1} \leqq \sum_{x \varepsilon G} v(x)^n.$$

Equality holds if and only if every $\tilde{\mathfrak{G}}_{n+1} : \tilde{\mathfrak{G}}_{n+1}$ double coset in $\mathfrak{G}_{n+1}$ is self inverse.

Specializing to $n = 2$ and applying Theorem 4 we obtain at once the inequality and the characterization of simply reducible groups which is Theorem 2 of Wigner's paper [6]. If in Theorem 6 we set $n = 1$ we find that $\sum_{x \varepsilon G} \zeta(x)^2$ is the order of $\mathfrak{G}$ multiplied by the number of self inverse classes in $\mathfrak{G}$. This is Wigner's Theorem 1.

It is natural to ask what equality means in terms of representation theory for values of n other than 2 and it turns out to be possible to give a complete answer.

THEOREM 8. *The following conditions on a finite group $\mathfrak{G}$ are equivalent*

(a) $\sum_{x \varepsilon G} \zeta(x)^2 = \sum_{x \varepsilon G} v(x)$.

(b) *Every class in $\mathfrak{G}$ is self inverse.*

(c) *For every representation L of $\mathfrak{G}$, L and $\bar{L}$ are equivalent.*

Proof. Let $\tilde{\mathfrak{G}}$ be the diagonal of $\mathfrak{G} \times \mathfrak{G}$. Let I be the identity representation of $\tilde{\mathfrak{G}}$. Then if L and M are irreducible representations of $\mathfrak{G}$, U^I contains $L \times M$ as a component just as many times as $L \otimes M$ contains I. But $L \times M$ contains I exactly once or not at all depending upon whether or not $M = \bar{L}$. Hence $U^I = \sum L \times \bar{L}$ where L ranges over all irreducible representations of $\mathfrak{G}$. Hence by Theorem 3 $\mathfrak{J}_A(\overline{U^I}, U^I) = 0$ if and only if $c(L \times \bar{L}) = 1$ for all L. But $c(L \times \bar{L}) = c(L)^2$. Thus $\mathfrak{J}_A(\overline{U^I}, U^I) = 0$ if and only if $L = \bar{L}$ for all L. It now follows from Theorem 3 that (b) and (c) are equivalent. That (a) and (b) are equivalent follows at once from Theorem 7.

It is to be remarked that the equivalence of (b) and (c) is a well known result.

THEOREM 9. *The following conditions on a finite group $\mathfrak{G}$ are equivalent.*

(a) *For some integer $n \geqq 3$, $\sum_{x \varepsilon G} \zeta(x)^{n+1} = \sum_{x \varepsilon G} v(x)^n$.*

(b) *For every positive integer* n, $\sum_{x \varepsilon G} \zeta(x)^{n+1} = \sum_{x \varepsilon G} v(x)^n$.

(c) *If* L *is any irreducible representation of* $\mathfrak{G}$ *then* $L \otimes \bar{L}$ *is the identity.*

(d) $\mathfrak{G}$ *is a direct product of groups of order two.*

Proof. It is obvious that (c) and (d) are equivalent and that (d) implies (a) and (b). It is also obvious that (b) implies (a). Now the condition: Every $\tilde{\mathfrak{G}}_{n+1} : \tilde{\mathfrak{G}}_{n+1}$ double coset in $\mathfrak{G}_{n+1}$ is self inverse can be reformulated to read: Given any n elements $x_1, x_2, \cdots, x_n$ in $\mathfrak{G}$ there exists $s \varepsilon \mathfrak{G}$ such that $sx_js^{-1} = x_j^{-1}$ for $j = 1, 2, \cdots, n$. Thus when the condition holds for any n it holds for all smaller n. Thus we need only prove that if (a) holds for $n = 3$ then (d) holds. But if (a) holds with $n = 3$ then by Theorem 7 and Theorem 3 we have $\mathfrak{J}_A(U^I, U^I) = 0$ where U^I is the representation of $\mathfrak{G} \times \mathfrak{G} \times \mathfrak{G} \times \mathfrak{G}$ induced by the identity representation of the diagonal $\tilde{\mathfrak{G}}_4$. Hence if L, M, N and K are arbitrary irreducible representations of $\mathfrak{G}$ then $L \times M \times N \times K$ occurs with multiplicity one or zero in U^I. Hence in particular $(L \otimes M) \otimes (\bar{L} \otimes \bar{M})$ contains the identity at most once. Hence $L \otimes M$ is irreducible for all irreducible L and M of $\mathfrak{G}$. Hence $L \otimes \bar{L}$ is irreducible for all L. But $L \otimes \bar{L}$ always contains the identity. Hence $L \otimes \bar{L}$ is the identity representation. Hence all irreducible representations of $\mathfrak{G}$ are one dimensional. Hence $\mathfrak{G}$ is Abelian and every element of its character group is of order 2. Since $\mathfrak{G}$ is isomorphic to its character group (d) follows and the theorem is proved.

5. Even and odd representations. Wigner calls an irreducible representation *even* if it occurs as a component in the reduction of $L \circledS L$ for some irreducible L with $c(L) = 1$ or in $L \circledA L$ for some irreducible L with $c(L) = -1$. He calls it *odd* if it occurs as a component in the reduction of $L \circledA L$ for some irreducible L with $c(L) = 1$ or[2] in $L \circledS L$ for some irreducible L with $c(L) = -1$. The last theorem of Wigner's paper (Theorem 3 of [6]) asserts that no representation of a simply reducible group can be both even and odd. In the language of intertwining numbers this theorem says that when $\mathfrak{G}$ is simply reducible and L and M are irreducible then $c(L)c(M) = 1$ implies that $\mathfrak{J}(L \circledS L, M \circledA M) = 0$ and $c(L)c(M) = -1$ implies that $\mathfrak{J}(L \circledS L, M \circledS M) = \mathfrak{J}(L \circledA L, M \circledA M) = 0$. On the other

[2] Actually this last phrase is omitted in Wigner's definition. For reasons of symmetry we believe that this omission must have been accidental. The theorem in question is stronger and still true when the definition is given as above.

hand when $\mathfrak{G}$ is simply reducible $L \otimes M$ is free of multiplicities and by Lemma 4 $c(V) = c(L)c(M)$ for all irreducible components V of $L \otimes M$. Thus if $c(L)c(M) = 1$ then $\mathfrak{I}_A(\overline{L \otimes M}, L \otimes M) = 0$ by the corollary to Lemam 5. By an obvious analogous argument if $c(L)c(M) = -1$ then $\mathfrak{I}_S(\overline{L \otimes M}, L \otimes M) = 0$. Thus Wigner's Theorem 3 is a consequence of the following lemma whose proof is an immediate consequence of the definitions.

LEMMA 7. *Let L and M be irreducible representations of an arbitrary finite group $\mathfrak{G}$. Then*

$$(L \circledS L) \otimes (M \circledA M) \subseteq (L \otimes M) \circledA (L \otimes M)$$
$$(L \circledS L) \otimes (M \circledS M) \subseteq (L \otimes M) \circledS (L \otimes M)$$
$$(L \circledA L) \otimes (M \circledA M) \subseteq (L \otimes M) \circledS (L \otimes M)$$

6. The conjugating representation of a group. Let $\mathfrak{H}$ be the set of all $\mathfrak{F}$ valued functions on the finite group $\mathfrak{G}$. For each $s \varepsilon \mathfrak{G}$ let A_s be the operator in $\mathfrak{H}$ such that $(A_s(f))(x) = f(s^{-1}xs)$. Then A is a representation of $\mathfrak{G}$ which Frame [2] has called the conjugating representation. We shall show that most of the results of Frame's paper follow at once from Theorem 1 of [4] and the first part of the proof of Theorem 8 of the present paper.

THEOREM 10. *Let $\mathfrak{G}$ be a finite group and let $\tilde{\mathfrak{G}}$ be the isomorphic replica furnished by the diagonal in $\mathfrak{G} \times \mathfrak{G}$. Let I be the identity representation of $\tilde{\mathfrak{G}}$. Then the following representations of $\tilde{\mathfrak{G}}$ are equivalent.*

(a) *The conjugating representation of $\tilde{\mathfrak{G}}$.*

(b) *The restriction to $\tilde{\mathfrak{G}}$ of the induced representation U^I of $\mathfrak{G} \times \mathfrak{G}$.*

(c) $\sum_{c \varepsilon \mathfrak{C}} P_c$ *where $\mathfrak{C}$ is the set of all classes in $\tilde{\mathfrak{G}}$ and P_c is the permutation representation of $\tilde{\mathfrak{G}}$ associated with the normalizer of any element in that class.*

(d) $\sum L \otimes \bar{L}$ *where L varies over all irreducible representations of $\tilde{\mathfrak{G}}$.*

Proof. $\mathfrak{H}(U^I)$ is the set of all $\mathfrak{F}$ valued functions f on $\mathfrak{G} \times \mathfrak{G}$ such that $f(tx, ty) = f(x, y)$ for all $t, x, y \varepsilon \mathfrak{G}$ and $(U^I{}_s(f))(x, y) = f(xs, ys)$. Let $(T(f))(z, z) = f(e, z)$ where e is the identity of $\mathfrak{G}$. Then T is one-to-one and linear from $\mathfrak{H}(U^I)$ to the set of all $\mathfrak{F}$ valued functions on $\tilde{\mathfrak{G}}$ and an obvious calculation shows that $s \to TU^I{}_sT^{-1}$ is the conjugating representation of $\tilde{\mathfrak{G}}$. Thus (a) and (b) are equivalent. Now by Theorem 1 of [4] U^I

restricted to $\tilde{\mathfrak{G}}$ is a sum over the double cosets of certain induced representations of $\tilde{\mathfrak{G}}$; the induced representation associated with the double coset containing x, y being that induced by the identity representation of $\tilde{\mathfrak{G}} \cap (x, y)^{-1}\tilde{\mathfrak{G}}(x, y)$. But this last subgroup is simply the set of all z, z such that $x^{-1}zx = y^{-1}zy$ or $(yx^{-1})z = z(yx^{-1})$; that is the normalizer of yx^{-1}. Since the classes in $\mathfrak{G}$ and the $\tilde{\mathfrak{G}} : \tilde{\mathfrak{G}}$ double cosets in $\mathfrak{G} \times \mathfrak{G}$ are in one-to-one correspondence in such a manner that x_1, y_1 and x_2, y_2 are in the same double coset if and only if $y_1x^{-1}{}_1$ and $y_2x^{-1}{}_2$ are in the same class, the equivalence of (b) and (c) follows at once. The first four sentences in the proof of Theorem 8 prove that $U^I = \sum L \times \bar{L}$ where the sum is over all irreducible representations L of $\mathfrak{G}$. Thus U^I restricted to $\tilde{\mathfrak{G}}$ is $= \sum L \otimes \bar{L}$; that is (b) and (d) are equivalent. This completes the proof of the theorem.

The equivalence of (a), (c) and (d) is established in Frame's paper ([2], Theorem 1). We remark with Frame that the equivalence of (a) and (d) furnishes a proof of the well known fact that the number of irreducible representations of a group is equal to the number of classes. Indeed in each sum every summand contains the identity exactly once.

We not also that if s is in the center of $\mathfrak{G}$ then $f(sxs^{-1}) = f(x)$ for all f and x so that the conjugating representation reduces to the identity on the center. Hence if the center of $\mathfrak{G}$ is non trivial no representation of the form $L \otimes \bar{L}$ where L is irreducible can be faithful. This is Frame's Theorem 2.

7. The infinite case. In part at least the results of the present paper may be extended to infinite dimensional unitary representations of locally compact topological groups. We have not yet actually written down the proof but there seems to be no difficulty in adapting the methods used in [5] in extending Theorem 2 of [4] to the infinite case and thus obtaining a corresponding extension of Theorem 1 of the present paper. The same remark applies to Theorem 2 and 2′ so long as we deal with strong intertwining numbers in the sense of [5] or consider only almost periodic representations. For compact groups at least then there should be no difficulty in establishing the obvious generalizations of Theorems 3, 4 and 10, the equivalence of (b) and (c) in Theorems 8 and that form of Theorem 9 in which the equation involving ζ and v are replaced by statements asserting the self inverseness of certain double cosets. For non compact groups the situation is more complicated and has not yet been thoroughly explored. We hope to give details in the compact case and do what can be done in more general cases in a subsequent paper.

HARVARD UNIVERSITY.

BIBLIOGRAPHY.

[1] Frame, J. S., "The double cosets of a finite group," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 458-467.

[2] ———, "On the reduction of the conjugating representation of a finite group," *Ibid.*, vol. 53 (1947), pp. 584-589.

[3] Frobenius, G. and Schur, I., "Über die reelen Darstellungen der endlichen Gruppen," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, vol. — (1906), pp. 186-208.

[4] Mackey, G. W., "On induced representations of groups," *American Journal of Mathematics*, vol. 73 (1951), pp. 576-592.

[5] ———, "Induced representations of locally compact groups I," *Annals of Mathematics*, vol. 55 (1952), pp. 101-139.

[6] Wigner, E. P., "On representations of certain finite groups," *American Journal of Mathematics*, vol. 63 (1941), pp. 57-63.

ON VALUES OMITTED BY UNIVALENT FUNCTIONS.*

By JAMES A. JENKINS.

1. Let S denote the family of functions $f(z)$ regular and univalent for $|z| < 1$ with the expansion $f(z) = z + a_2 z^2 + \cdots$ about $z = 0$. It is well known that $f(z) \varepsilon S$ assumes for $|z| < 1$ all values w with $|w| < \frac{1}{4}$. Further it can omit just one value w with $|w| = \frac{1}{4}$. The latter can occur only for certain well known and explicitly given functions. On the other hand for $\rho \geqq 1$ there exists a function $f(z) \varepsilon S$ omitting the entire circle $|w| = \rho$ (for example $f(z) \equiv z$). It seems natural to inquire how many values can be omitted on a circle $|w| = r$, $\frac{1}{4} < r < 1$. Let $L(f, r)$ denote the length of the set of values on $|w| = r$ not covered by values of $f(z) \varepsilon S$ for $|z| < 1$. In this paper we will give the precise upper bound for $L(f, r)$ for $f \varepsilon S$ and for each value of r in $\frac{1}{4} < r < 1$. We will then apply this result to improve a result of A. W. Goodman [1].

2. We begin by constructing an explicit conformal mapping. Regard in the ζ-plane the domain D bounded by $|\zeta| = 1$ together with the portion of the real axis $1 \leqq \zeta \leqq \infty$ and distinguish the point $\zeta_0 = -\rho$ $(\rho > 1)$ within this domain. The function $\eta = \zeta + \zeta^{-1} + 2$ maps this domain on the η-plane slit along the positive real axis. ζ_0 goes into $\eta_0 = -(\rho + \rho^{-1} - 2)$. The function $W = \eta^{\frac{1}{2}}$ (taking the positive determination on the upper side of the positive real axis) maps the preceding domain on the upper half W-plane, η_0 going into $W_0 = i(\rho^{\frac{1}{2}} - \rho^{-\frac{1}{2}})$. Finally $z = -(W - W_0)/(W - \bar{W}_0)$ maps the latter domain on $|z| < 1$, W_0 going into $z = 0$. An elementary calculation shows that $d\zeta/dz\,|_{z=0} = 4\rho(\rho - 1)/(\rho + 1)$. On the other hand the function $Z = \rho(\zeta^2 + \rho\zeta)/(\rho\zeta + 1)$ maps D on a domain D^* bounded by an arc on $|Z| = \rho$ placed symmetrically with respect to the positive real axis together with the portion $\rho \leqq Z \leqq \infty$ of the latter. The point ζ_0 goes into $Z = 0$. By an elementary calculation $dZ/d\zeta\,|_{\zeta=\zeta_0} = \rho^2/(\rho^2 - 1)$. As a function of z, Z maps $|z| < 1$ on the domain D^*, and $dZ/dz\,|_{z=0} = 4\rho^3/(\rho + 1)^2$.

Thus the function $w = (\rho + 1)^2 Z/4\rho^3$ belongs to S. It maps $|z| < 1$ on a domain $\tilde{D}(r)$ bounded by an arc on the circle $|w| = r = (\rho + 1)^2/(2\rho)^2$

* Received October 24, 1952.

placed symmetrically with respect to the positive real axis, together with the portion $r \leqq w \leqq \infty$ of the latter. As ρ takes the values in $\infty > \rho > 1$, r takes the values in $\frac{1}{4} < r < 1$.

We wish now to determine the length of the arc on $|w| = r$. The end points of the corresponding arc in the Z-plane are the images of the points on $|\zeta| = 1$ where $dZ/d\zeta = 0$ i. e. the solutions of $\rho\zeta^2 + 2\zeta + \rho = 0$. These are the points $\zeta = (-1 \pm i(\rho^2 - 1)^{\frac{1}{2}})/\rho$ and their images are $Z = (\rho^2 - 2 \pm 2i(\rho^2 - 1)^{\frac{1}{2}})/\rho$. The angle subtended at the origin by this arc is $\theta = 2\cos^{-1}(1 - 2/\rho^2)$ (the principal branch of $\cos^{-1}$ being used). Now $\rho = (2r^{\frac{1}{2}} - 1)^{-1}$ thus $\theta = 2\cos^{-1}(8r^{\frac{1}{2}} - 8r - 1)$. Hence the length of the arc is $2r\cos^{-1}(8r^{\frac{1}{2}} - 8r - 1)$. We observe this tends to 0 as r tends to $\frac{1}{4}$, and tends to 2π as r tends to 1.

3. THEOREM. *For* $\frac{1}{4} < r < 1$, $L(f, r) \leqq 2r\cos^{-1}(8r^{\frac{1}{2}} - 8r - 1)$.

For $f \varepsilon S$ let $D(f)$ be the image domain of $|z| < 1$ under f. Let $D^*(f)$ be the domain obtained from $D(f)$ by circular symmetrization in the following manner: for all s, $0 < s < \infty$ if $D(f) \cap \{|w| = s\}$ consists of the whole circumference $|w| = s$, $D^*(f) \cap \{|w| = s\}$ shall do the same; otherwise $D^*(f) \cap \{|w| = s\}$ shall consist of a single arc on $|w| = s$ of length equal to that of $D(f) \cap \{|w| = s\}$ and centred at the point $w = -s$. It is well known that we obtain in this way a domain $D^*(f)$ whose inner conformal radius with respect to $w = 0$ is at least 1. Unfortunately there seems no reference where this result is treated with complete precision. It seems to be most readily established by using the methods of Pólya and Szegö [3, 4]. A rough sketch of a proof appears in the report of Hayman [2]. For a careful treatment of similar problems by a different method see [5].

Now supose that we had $L(f, r) > 2r\cos^{-1}(8r^{\frac{1}{2}} - 8r - 1)$ for some $f \varepsilon S$ and some r, $\frac{1}{4} < r < 1$. Then $D^*(f)$ would be a proper subdomain of the domain $D(r)$ of § 2. Since the latter has inner conformal radius 1 with respect to the origin, we would be led to a contradiction. This proves the theorem.

4. A few years ago Goodman [1] proved that the greatest lower bound A of $A_f = \text{area } D(f) \cap \{|w| < 1\}$ for $f \varepsilon S$ satisfies $.5000\pi \leqq A < .7728\pi$. The preceding theorem enables us to improve the first bound to $.5387\pi < A$. Indeed the area of the part of $|w| < 1$ omitted by f for $|z| < 1$ cannot exceed

$$\int_{\frac{1}{4}}^{1} 2r \cos^{-1}(8r^{\frac{1}{2}} - 8r - 1)\, dr.$$

Integrating this explicitly by elementary means we find its value $.4613\pi$ to that degree of accuracy. Thus the area covered is at least $.5387\pi$. It is clear that this is still apreciably less than the true greatest lower bound.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] A. W. Goodman, "Note on regions omitted by univalent functions," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 363-369.

[2] W. K. Hayman, "Symmetrisation in the theory of function," *Technical Report No.* 11, Contract N6-ORI-106, Task order 5 (NR-043-992) O.N.R., Washington.

[3] G. Pólya, "Sur la symétrisation circulaire," *Comptes rendus de l'Académie des Sciences*, vol. 230 (1950), pp. 25-27.

[4] ——— and G. Szegö, "Inequalities on the capacity of a condenser," *American Journal of Mathematics*, vol. 67 (1945), pp. 1-32.

[5] V. Wolontis, "Properties of conformal invariants," *ibid.*, vol. 74 (1952), pp. 587-606.

GROUPES DE LIE ET PUISSANCES RÉDUITES DE STEENROD.*

Par A. Borel et J.-P. Serre.

Introduction. N. E. Steenrod a défini de nouvelles opérations cohomologiques, les puissances réduites, qui généralisent ses i-carrés. Nous nous proposons ici d'étudier ces opérations dans la cohomologie mod p, (p premier), des groupes de Lie et de leurs espaces classifiants, et d'appliquer les résultats obtenus à divers problèmes.

Pour la commodité du lecteur, nous avons rappelé dans la première partie tous les principaux résultats sur les groupes de Lie et leurs espaces classifiants dont nous avons à faire usage par la suite, en les complétant du reste sur quelques points. La deuxième partie est consacrée aux puissances réduites, dont nous indiquons les propriétés au No. 7, sans en répéter la définition explicite, qui n'interviendra pas ici; nous calculons ensuite ces opérations dans les espaces projectifs complexes et quaternioniens, ce qui permet d'obtenir quelques renseignements sur les groupes d'homotopie des sphères, par la méthode de Steenrod.

Dans la troisième partie, nous combinons les résultats de I et II pour étudier les puissances réduites dans les algèbres de cohomologie $H^*(G, Z_p)$ et $H^*(B_G, Z_p)$ d'un groupe de Lie compact connexe G et d'un espace B_G classifiant pour G, *lorsque G et son quotient G/T par un tore maximal sont sans p-torsion.*[1] Dans ce cas en effet, $H^*(B_G, Z_p)$ s'identifie à une sous-algèbre de $H^*(B_T, Z_p)$; or $H^*(B_T, Z_p)$ est engendrée par ses éléments de degré deux, (B_T peut même, si l'on veut, être envisagé comme produit d'espaces projectifs complexes), et les puissances réduites y sont donc connues d'après II. Cela détermine en principe les puissances réduites dans $H^*(B_G, Z_p)$, et par conséquent aussi dans $H^*(G, Z_p)$ car, sous les hypothèses faites, $H^*(B_G, Z_p)$ est une algèbre de polynômes dont les générateurs sont images par transgression de générateurs de $H^*(G, Z_p)$, (qui est une algèbre extérieure), et la transgression commute aux puissances réduites. Cette méthode générale est ensuite appliquée aux groupes unitaire $\boldsymbol{U}(n)$ et unitaire symplectique $\boldsymbol{Sp}(n)$ pour p (premier) quelconque, au groupe orthogonal

* Received February 22, 1953.

[1] On dit qu'un espace a de la p-torsion (p premier), si l'un de ses groupes d'homologie entière a un coefficient de torsion divisible par p.

$\boldsymbol{SO}(n)$ pour $p \neq 2$, et à leurs espaces classifiants, ce qui donne, entre autres, des résultats sur les classes de Chern.

La quatrième partie est consacrée aux applications. Nous montrons dans le No. 15 la non-existence de structures presque-complexes sur $\boldsymbol{S}_n$, $(n \geq 8)$, et d'un type d'algèbres à division de dimension > 8, et dans le No. 17, la non-existence de sections dans certaines fibrations, (par exemple $\boldsymbol{U}(n)/\boldsymbol{U}(n-1) = \boldsymbol{S}_{2n-1}$) ; on en déduit quelques renseignements sur les groupes d'homotopie des groupes classiques, auxquels nous consacrons également les Nos. 18 et 19. Enfin, le No. 20 donne des conditions nécessaires pour l'existence de sections dans des fibrations où espace, base et fibre sont des variétés de Stiefel complexes.

I. Espaces fibrés à groupe structural de Lie.

1. Espaces universels et espaces classifiants pour un groupe de Lie. Soit G un groupe de Lie compact; rappelons que l'on appelle *espace universel pour* G jusqu'à la dimension n un espace E, fibré principal de groupe structural G, tel que $\pi_i(G) = 0$ pour $0 \leq i \leq n$, ([14], § 19) ; sa base $B = E/G$ est dite *espace classifiant pour* G et pour la dimension n, elle permet en effet de classer tous les espaces fibrés principaux de groupe structural G ayant comme base un polyèdre X donné de dimension n, (une classe d'espaces fibrés correspondant biunivoquement à une classe d'applications homotopes de X dans B, *voir* [14], § 19).

On sait, ([14], 19. 7), que l'on peut trouver pour tout G et pour tout n des espaces universels qui sont, ainsi que leurs bases, des variétés analytiques compactes, donc des polyèdres finis; il est souvent commode de les considérer pour n arbitrairement grand et pour éviter d'avoir à préciser cet entier, on peut introduire les notions d'espace universel et d'espace classifiant pour G, (sous-entendu: pour tout n), de la façon suivante:

Soit $E_1, E_2, \cdots$ une suite d'espaces universels pour G et pour des dimensions $n_1 < n_2 < \cdots$; il est clair que l'on peut la choisir telle que E_i et $B_i = E_i/G$ soient des polyèdres finis et qu'il existe un homéomorphisme f_i de E_i dans E_{i+1} commutant avec les opérations de G, $(i = 1, 2, \cdots)$. La limite inductive E des E_i est alors un espace fibré principal de groupe structural G, dont tous les groupes d'homotopie sont nuls; E sera dit universel pour G, sa base B, qui est limite inductive des espaces B_i, sera dite espace classifiant pour G. On déduit immédiatement du théorème de classification que deux espaces universels, ou deux espaces classifiants, ont même type d'homotopie; il n'y a donc pas d'inconvénient, tant que l'on ne s'intéresse qu'à

des questions homologiques ou homotopiques, à désigner par E_G, resp. B_G, l'un quelconque de ces espaces; en particulier, on pourra parler des groupes d'homologie singulière ou de cohomologie singulière de l'espace B_G; on a d'ailleurs, Γ étant un groupe de coefficients:

$$H_p(B_G, \Gamma) \approx H_p(B_i, \Gamma), \qquad H^p(B_G, \Gamma) \approx H^p(B_i, \Gamma) \qquad \text{si } n_i > p,$$

(les isomorphismes en cohomologie étant bien entendu compatibles avec le cup-produit si Γ est un anneau) ; les groupes $H^p(B_G, \Gamma)$ sont donc isomorphes aux groupes $H^p(B_i, \Gamma)$ lorsque i est assez grand, on retrouve ainsi les conventions de [2], § 18.[2]

Nous avons donc fait correspondre à tout groupe de Lie compact G un espace, ou plutôt une classe d'espaces ayant le même type d'homotopie, B_G. On peut de plus associer à tout homomorphisme $f: H \to G$ d'un groupe de Lie compact dans un autre une classe d'applications homotopes $\rho(f): B_H \to B_G$.

Soit en effet E_H un espace universel pour H; par extension du groupe structural de H à G au moyen de f, E_H définit un espace E'_H principal pour G et de base B_H; cet espace est le quotient $(E_H, G)_H$ du produit $E_H \times G$ par la relation d'équivalence $(x, g) \approx (x \cdot h, f(h) \cdot g)$. Comme la base de E'_H est limite inductive de polyèdres, le théorème de classification s'applique, et il existe une classe d'applications homotopes $\rho(f)$ de B_H dans B_G telle que E'_H soit l'image réciproque de E_G par les $\rho(f)$; cela définit les $\rho(f)$.

Bien entendu, $\rho(f)$ est homotope à l'identité si f est l'identité, et les $\rho(f)$ vérifient la propriété de transitivité $\rho(f \circ g) = \rho(f) \circ \rho(g)$. On peut donc dire que la correspondance $G \to B_G$ est un *foncteur covariant* de G.

Le cas particulier le plus important de la notion précédente est celui où f est l'application identique d'un sous-groupe fermé H d'un groupe G dans le groupe G lui-même; on désigne alors $\rho(f)$ par $\rho(H, G)$, conformément aux notations de [2], § 21. Dans ce cas on peut choisir $\rho(f)$ de telle sorte que *$\rho(f): B_H \to B_G$ définisse B_H comme espace fibré de base B_G et de fibre l'espace homogène G/H.* En effet, puisque H est plongé dans G, il opère sur E_G et E_G, muni de ces opérateurs, est un espace universel pour H, que l'on peut prendre comme E_H. On aura alors $B_H = E_G/H$, $B_G = E_G/G$ et l'application $\rho(f)$ n'est autre que la projection canonique de E_G/H sur E_G/G; elle définit bien B_H comme espace fibré de fibre G/H et de base B_G.

Dans le cas général, on peut, au moins au point de vue homologique ou

---

[2] Si G est discret (cas que nous n'avons pas exclu), les groupes $H_p(B_G, \Gamma)$ et $H^p(B_G, \Gamma)$ ne sont autres que les groupes d'homologie et de cohomologie de G au sens de Hopf-Eilenberg-MacLane-Eckmann.

homotopique, envisager aussi $\rho(f)$ comme projection d'un espace fibré. Soit en effet $X = (E'_H, E_H)_H$ le quotient de $E'_H \times E_H$ par la relation d'équivalence $(x, y) \approx (x \cdot h, y \cdot h)$, il admet deux fibrations, l'une de fibre E'_H et de base B_H, l'autre de fibre E_H et de base B_G; notons α et β les projections correspondantes. Comme E_H est acyclique, β définit un isomorphisme β_* des groupes d'homologie (ou d'homotopie) de X sur ceux de B_H. D'autre part un homomorphisme de E'_H dans E_H définit de façon évidente une section s: $B_H \rightarrow X$, qui composée avec α redonne $\rho(f)$; ainsi, une fois les groupes d'homologie de B_H identifiés à ceux de X par β_*^{-1}, l'homomorphisme $\rho_*(f)$ devient l'homomorphisme α_* induit par la projection α.

En particulier, si f est la projection de H sur son quotient H/N par un sous-groupe invariant fermé N, l'espace E'_H est visiblement un espace B_N et $\rho_*(f)$ *s'identifie à l'homomorphisme induit par la projection d'un espace fibré ayant même homologie que B_H, de fibre B_N et de base $B_{H/N}$.*

Note. Soit E un espace fibré de groupe structural H dont la base X est un polyèdre fini; E est donc bien défini par une classe d'applications homotopes $\zeta: X \rightarrow B_H$; en les composant avec les applications $\rho(f): B_H \rightarrow B_G$, on définit donc un espace fibré de base X et de groupe structural G; en outre, d'après la construction même de $\rho(f)$, cet espace est celui que l'on obtient à partir de E *en étendant le groupe structural de H à G au moyen de f.*

C'est sous cette forme que l'on trouvera étudiée, dans des cas particuliers, l'application $\rho(f)$, notamment par Wu ([18], [19]). On voit également que pour qu'un espace fibré de base X, de groupe G, défini par $\zeta: X \rightarrow B_G$ puisse être obtenu à partir d'un espace fibré de groupe H par extension du groupe structural il faut et il suffit que ζ puisse se "factoriser" par $\rho(f)$. Si l'on connait les algèbres de cohomologie $H^*(B_H)$ et $H^*(B_G)$, ainsi que $\rho^*(f)$: $H^*(B_G) \rightarrow H^*(B_H)$, on tire de là des conditions cohomologiques nécessaires pour que l'on puisse restreindre le groupe structural de G à H. C'est là la méthode suivie par Wu [18] pour étudier les structures presque complexes, (f étant alors l'inclusion de $\boldsymbol{U}(n)$ dans $\boldsymbol{SO}(2n)$).

2. Cohomologie des groupes de Lie et de leurs espaces classifiants. Soit G un groupe de Lie compact connexe de rang l, (rappelons que le rang est la dimension commune des tores maximaux de G). D'après un théorème classique de Hopf, l'algèbre de cohomologie de G relativement à un corps de caractéristique zéro est une algèbre extérieure engendrée par l éléments de degrés impairs. Ce résultat vaut encore pour $H^*(G, Z_p)$, (p premier, Z_p corps des entires modulo p), lorsque G est sans p-torsion,[1] ([2], Proposition 7. 2);

de même, si G n'a pas de torsion, $H^*(G, Z)$ est l'algèbre extérieure d'un groupe abélien libre ayant l générateurs de degrés impairs.

Dans le cas où G est sans p-torsion, on peut établir des relations très précises entre $H^*(G, Z_p)$ et $H^*(B_G, Z_p)$ en utilisant la transgression. Rappelons que la transgression dans un espace fibré E, de base B et de fibre F, relativement à un groupe de coefficients Γ, est en dimension s, $(s = 0, 1, 2, \cdots)$, un homomorphisme:

$$(2.1) \qquad \tau : T^s(F, \Gamma) \to H^{s+1}(B, \Gamma)/L^{s+1}(B, \Gamma)$$

d'un sous-groupe $T^s(F, \Gamma)$ de $H^s(F, \Gamma)$ dans un quotient de $H^{s+1}(B, \Gamma)$. L'homomorphisme τ est le composé $q^{*-1} \circ \delta$, où δ désigne l'homomorphisme de cobord qui applique $H^s(F, \Gamma)$ dans $H^{s+1}(E, F; \Gamma)$, et où q^* est le produit de l'isomorphisme de $H^{s+1}(B, \Gamma)$ sur $H^{s+1}(B, b; \Gamma)$, (b projection de F), par l'homomorphisme $p^* : H^{s+1}(B, b; \Gamma) \to H^{s+1}(E, F; \Gamma)$ transposé de la projection (*voir* [2], § 5; [11], pp. 434, 457). En particulier nous noterons $T^s(G, \Gamma)$ l'ensemble des éléments de $H^s(G, \Gamma)$ transgressifs dans un espace universel E_G, et qui seront dits être *universellement transgressifs.*

Ces définitions étant posées, on peut exprimer ainsi les propriétés de la fibration de E_G par G, base B_G qui sont établies dans [2], (Théorèmes 13. 1, 19. 1):

2. 2. *Soit p un nombre premier, et supposons G sans p-torsion. Alors $H^*(G, Z_p)$ possède un système de générateurs $h_1, \cdots, h_l$ de degrés impairs qui forment une base du sous-espace $T(G, Z_p)$ de $H^*(G, Z_p)$ engendré par les éléments universellement transgressifs.*

2. 3. *Le sous-espace $L^{s+1}(B_G, Z_p)$ de $H^{s+1}(B_G, Z_p)$,* (notations de 2. 1), *est égal au sous-espace des éléments décomposables de $H^{s+1}(B_G, Z_p)$,* (c'est à dire au sous-espace engendré par les produits d'éléments de degrés $< s + 1$).

Nous désignerons par $D^j(B_G, \Gamma)$, ou par D^j si cela ne prête pas à confusion, l'ensemble des éléments décomposables de $H^j(B_G, \Gamma)$ et par $D(B_G, \Gamma)$, ou par D, la somme directe des D^j.

2. 4. *Soient τ la transgression dans E_G et $y_i \in H^*(B_G, Z_p)$ un représentant de $\tau(h_i)$, $(i = 1, \cdots, l)$. Alors $H^*(B_G, Z_p)$ est identique à l'algèbre des polynômes admettant les y_i comme générateurs.*

On a ici $\tau(h_i) = y_i \bmod D$ et $H^*(B_G, Z_p)$ est une algèbre de polynômes à l générateurs dont les degrés sont égaux aux degrés des h_i augmentés d'une unité (et sont par conséquent pairs).

Les résultats 2. 2, 2. 3, 2. 4 restent valables si l'on remplace partout Z_p par un corps de caractéristique zéro, sans hypothèse sur G, ou encore si l'on substitue Z à Z_p lorsque G n'a pas de torsion.

3. Relations entre $H^*(B_G, Z_p)$ et le groupe de Weyl de G. Soient T un tore maximal du groupe de Lie compact connexe G de rang l, N le normalisateur de T dans G, et Φ le groupe de Weyl de G, c'est à dire le quotient N/T. On sait que Φ est un groupe fini. D'après les résultats du No. 2, la transgression établit un isomorphisme de $H^1(T, Z)$ sur $H^2(B_T, Z)$, et $H^*(B_T, Z)$ est l'algèbre symétrique libre engendrée par $H^2(B_T, Z)$; toute base $(\xi_1, \cdots, \xi_l)$ de $H^1(T, Z)$ définit donc par transgression un système de générateurs indépendants $(x_1, \cdots, x_l)$ de $H^*(B_T, Z)$ de degrés égaux à deux

Puisque N est une extension de T par Φ, le groupe Φ opère canoniquement sur T, et de ce fait sur $H^1(T, Z)$ donc aussi sur $H^*(B_T, Z)$. On notera I_G la sous-algèbre des éléments de $H^*(B_T, Z)$ invariants par Φ. Il est clair que si $x \in H^*(B_T, Z)$ est tel que $n \cdot x \in I_G$, (n entier), on a aussi $x \in I_G$; par conséquent I_G est *facteur direct* de $H^*(B_T, Z)$ pour la structure de groupe abélien, ce qui permet de considérer $I_G \otimes Z_p$ comme plongé dans

$$H^*(B_T, Z) \otimes Z_p \approx H^*(B_T, Z_p).$$

Cela étant posé, on a ([2], Prop. 29. 2):

3. 1. *Si G et G/T sont sans torsion, l'homomorphisme*

$$\rho^*(T, G): H^*(B_G, Z) \to H^*(B_T, Z)$$

est biunivoque et son image est I_G.

3. 2. *Soit p un nombre premier. Si G et G/T sont sans p-torsion, il en est de même de B_G, l'homomorphisme:*

$$\rho^*(T, G): H^*(B_G, Z_p) \to H^*(B_T, Z_p)$$

est biunivoque, et son image est $I_G \otimes Z_p$, (plongé dans $H^*(B_T, Z_p)$ comme il a été dit plus haut).

Lorsqu'on sera dans les hypothèses de 3. 1, (resp. 3. 2), on identifiera $H^*(B_G, Z)$, (resp. $H^*(B_T, Z_p)$), à I_G, (resp. $I_G \otimes Z_p$); cela permettra comme on le verra de ramener beaucoup de questions portant sur B_G aux questions analogues sur B_T, qui est plus simple à étudier. Soient par exemple H et G deux groupes de Lie vérifiant 3. 2, $f: H \to G$ un homomorphisme et cherchons à déterminer $\rho^*(f): H^*(B_G, Z_p) \to H^*(B_H, Z_p)$. Soient T' un tore maximal

de H, T un tore maximal de G contenant $f(T')$, et $g: T' \to T$ la restriction de f à T'. Elle définit un homomorphisme $\rho^*(g): H^*(B_T, Z_p) \to H^*(B_{T'}, Z_p)$ dont la restriction à $I_G \otimes Z_p$ est évidemment $\rho^*(f)$; pour connaître $\rho^*(f)$, il suffit donc de connaître $\rho^*(g)$, ce qui est très facile: Soient $\xi_1, \cdots, \xi_r$, (resp. $\xi'_1, \cdots, \xi'_s$), une base de $H^1(T, Z_p)$, (resp. de $H^1(T', Z_p)$. On peut écrire:

$$g^*(\xi_i) = \Sigma\, n_{ij}\xi'_j, \qquad (n_{ij} \in Z_p),$$

et, si x_i, resp. x'_i, est image par transgression de ξ_i, resp. ξ'_i, un polynôme $Q(x_1, \cdots, x_r)$ en les x_i a comme image par $\rho^*(g)$ le polynôme

$$Q(\Sigma\, n_{1j}x'_j, \cdots, \Sigma\, n_{rj}x'_j),$$

(*voir* pour plus de détails [2], § 28, 31 où cette méthode est appliquée dans le cas où f est biunivoque).

Un cas particulier important est celui où f est l'inclusion d'un sous-groupe H de G dans G, *H et G ayant même rang.* On a alors $T = T'$, et $\rho^*(g)$ est l'identité; tenant compte des identifications faites plus haut on voit que *$H^*(B_G, Z_p)$ s'identifie à une sous-algèbre de $H^*(B_H, Z_p)$, elle-même identifiée à une sous-algèbre de $H^*(B_T, Z_p)$.*

Conditions d'application des résultats précédents. Les hypothèses 3. 1 et 3. 2 sont fréquemment vérifiées, en effet:

3. 3. *Si l'algèbre de Lie de G ne contient aucun facteur isomorphe à $\boldsymbol{E}_6$, $\boldsymbol{E}_7$, ou $\boldsymbol{E}_8$, G/T est sans torsion* ([2], Prop. 29. 1).

3. 4. *Pour tout n les groupes classiques $\boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$ sont sans torsion* ([2], Prop. 9. 1).

3. 5. *Pour tout n et pour tout nombre premier p impair, $\boldsymbol{SO}(n)$ est sans p-torsion,* ([2], Prop. 10. 4).

Ainsi, *à l'exception de $\boldsymbol{SO}(n)$ pour $p = 2$, tout groupe classique est justiciable de* 3. 2. Dans le cas de $\boldsymbol{SO}(n)$, $p = 2$, il convient de remplacer les tores maximaux par des sous-groupes abéliens maximaux de type $(2, 2, \cdots, 2)$, (*voir* [3]).

4. Cas particulier: le groupe unitaire $\boldsymbol{U}(n)$. Soit $G = \boldsymbol{U}(n)$ le groupe des matrices complexes unitaires à n lignes et n colonnes; les matrices diagonales en constituent un tore maximal $\boldsymbol{T}$, $\boldsymbol{U}(n)$ est donc de rang n. Désignant par $\exp(2i\pi\zeta_1), \cdots, \exp(2i\pi\zeta_n)$ les valeurs propres d'une matrice

diagonale, nous prendrons les éléments $\xi_i = d\zeta_i$ comme base de $H^1(\boldsymbol{T}, Z)$ et noterons x_i l'image par transgression de ξ_i dans $H^2(B_{\boldsymbol{T}}, Z)$.

Le normalisateur $\boldsymbol{N}$ de $\boldsymbol{T}$ dans $\boldsymbol{U}(n)$, est l'ensemble des matrices *monomiales* (c'est à dire produits d'une matrice diagonale par une matrice de permutation), et $\Phi = \boldsymbol{N}/\boldsymbol{T}$ est le groupe des permutations des ξ_i ou des x_i; par conséquent I_G est l'algèbre des polynômes symétriques en les x_i; de même $I_{\boldsymbol{U}(n)} \otimes Z_p$ est l'ensemble des polynômes symétriques à coefficients dans Z_p.

Soit C_{2i} la i-ème fonction symétrique élémentaire $\Sigma x_1 \cdots x_i$ [3] c'est donc un élément de $I_G = H^*(B_{\boldsymbol{U}(n)}, Z)$ et de plus I_G est identique à l'algèbre des polynômes en les C_{2i}; comparant cela avec les résultats du No. 2, on voit que $H^*(\boldsymbol{U}(n), Z)$ contient des éléments bien déterminés h_i, $(1 \leq i \leq n)$, de degré $2i - 1$, tels que $\tau(h_i) = C_{2i} \operatorname{mod}(C_2, \cdots, C_{2i-2})$, τ étant la transgression dans $E_{\boldsymbol{U}(n)}$; en outre $H^*(\boldsymbol{U}(n), Z)$ est l'algèbre extérieure engendrée par les h_i.

Il s'impose de comparer les résultats précédents avec ceux qu'obtient S. S. Chern dans [6]. Au moyen des grassmanniennes complexes et des symboles de Schubert, Chern y définit des éléments $c_{2i} \varepsilon H^{2i}(B_{\boldsymbol{U}(n)}, Z)$ et montre que $H^*(B_{\boldsymbol{U}(n)}, Z)$ est l'algèbre des polynômes admettant les c_{2i}, $(1 \leq i \leq n)$, comme générateurs; ainsi les classes C_{2i} ont les mêmes propriétés que les classes de Chern c_{2i}; en fait on a:

PROPOSITION 4. 1. *Les classes* $C_{2i} = \Sigma x_1 \cdots x_i$ *coïncident avec les classes de Chern* c_{2i} $(i = 1, \cdots, n)$.

Pour éviter toute confusion, nous n'identifierons pas dans cette démonstration $I_{\boldsymbol{U}(n)}$ et $H^*(B_{\boldsymbol{U}(n)}, Z)$. Nous devons donc montrer que *l'homomorphisme* $\rho^*(\boldsymbol{T}, \boldsymbol{U}(n)) : H^*(B_{\boldsymbol{U}(n)}, Z) \to H^*(B_{\boldsymbol{T}}, Z)$ *applique* c_{2i} *sur* C_{2i}.

Nous établirons la Prop. 4. 1 par récurrence sur n en utilisant la *formule de dualité* des classes de Chern (4. 3), (*voir* [7], [19]). Pour $n = 1$, $\boldsymbol{U}(1) = \boldsymbol{T}$ et la proposition est évidente car C_2 et c_2 sont toutes deux images par transgression de la classe $\xi = d\zeta$ de $H^1(\boldsymbol{T}, Z)$. Nous supposons maintenant la proposition démontrée pour $\boldsymbol{U}(m)$ si $m < n$; comme nous aurons à considérer plusieurs valeurs de n, il sera commode de distinguer par un indice n, que nous placerons en haut à gauche, ce qui est relatif à $\boldsymbol{U}(n)$; ainsi nous parlerons de ${}^n\xi_i$, ${}^nx_i$, ${}^nC_{2i}$, ${}^nc_{2i}$, etc.

Soient $p, q > 0$ tels que $p + q = n$, f l'inclusion canonique de $\boldsymbol{U}(p) \times \boldsymbol{U}(q)$ dans $\boldsymbol{U}(n)$, $\boldsymbol{T}^p$, $\boldsymbol{T}^q$ et $\boldsymbol{T}^n = \boldsymbol{T}^p \times \boldsymbol{T}^q$ des tores maximaux de $\boldsymbol{U}(p)$, $\boldsymbol{U}(q)$ et

[3] Dans tout ce travail, nous notons un polynôme symétrique par son terme initial précédé du signe Σ.

$\boldsymbol{U}(n)$; la restriction g de f à $\boldsymbol{T}^p \times \boldsymbol{T}^q$ est donc l'identité. Le diagramme commutatif

$$\begin{array}{ccc} \boldsymbol{T}^p \times \boldsymbol{T}^q & \xrightarrow{g} & \boldsymbol{T}^n \\ \downarrow & & \downarrow \\ \boldsymbol{U}(p) \times \boldsymbol{U}(q) & \xrightarrow{f} & \boldsymbol{U}(n) \end{array}$$

où les flèches sont des inclusions, donne lieu au diagramme commutatif

$$\begin{array}{ccc} H^*(B_{\boldsymbol{T}^p}, Z) \otimes H^*(B_{\boldsymbol{T}^q}, Z) & \xleftarrow{\alpha} & H^*(B_{\boldsymbol{T}^n}, Z) \\ \uparrow \beta & & \uparrow \delta \\ H^*(B_{\boldsymbol{U}(p)}, Z) \otimes H^*(B_{\boldsymbol{U}(q)}, Z) & \xleftarrow{\gamma} & H^*(B_{\boldsymbol{U}(n)}, Z) \end{array}$$

où, dans les notations du No. 1:

$$\alpha = \rho^*(g) = \rho^*(\boldsymbol{T}^p \times \boldsymbol{T}^q, \boldsymbol{T}^n), \qquad \beta = \rho^*(\boldsymbol{T}^p \times \boldsymbol{T}^q, \boldsymbol{U}(p) \times \boldsymbol{U}(q)),$$
$$\gamma = \rho^*(\boldsymbol{U}(p) \times \boldsymbol{U}(q), \boldsymbol{U}(n)) = \rho^*(f) \text{ et enfin } \delta = \rho^*(\boldsymbol{T}^n, \boldsymbol{U}(n)).$$

Si l'on prend dans $H^1(\boldsymbol{T}^p, Z)$, $H^1(\boldsymbol{T}^q, Z$ et $H^1(\boldsymbol{T}^n, Z)$ les bases $({}^p\xi_j)$, $({}^q\xi_k)$, $({}^n\xi_i)$ indiquées au début de ce No, il est clair que g^* est définie par

$$g^*({}^n\xi_i) = {}^p\xi_i \quad (i \leq p), \qquad g^*({}^n\xi_{p+i}) = {}^q\xi_i, \quad (1 \leq i \leq q),$$

par conséquent, α est défini par

$$\alpha({}^n x_i) = {}^p x_i \otimes 1 \quad (1 \leq i \leq p); \qquad \alpha({}^n x_{p+j}) = 1 \otimes {}^q x_j \quad (1 \leq j \leq q),$$

et il en résulte évidemment que

$$\alpha({}^n C_{2i}) = \Sigma_{j+k=i}\, {}^p C_{2j} \otimes {}^q C_{2k} \tag{4. 2}$$

en posant bien entendu ${}^p C_{2j} = 0$ si $j > p$, ${}^q C_{2k} = 0$ si $k > q$; en faisant une convention analogue pour les classes de Chern, on peut écrire la formule de dualité:

$$\gamma({}^n c_{2i}) = \Sigma_{j+k=i}\, {}^p c_{2j} \otimes {}^q c_{2k} \tag{4. 3}$$

mais l'homomorphisme β est le produit tensoriel des homomorphismes $\rho^*(\boldsymbol{T}^p, \boldsymbol{U}(p))$ et $\rho^*(\boldsymbol{T}^q, \boldsymbol{U}(q))$, donc, vu l'hypothèse d'induction, on obtient:

$$\beta \circ \gamma({}^n c_{2i}) = \Sigma_{j+k=i}\, {}^p C_{2j} \otimes {}^q C_{2k} \tag{4. 4}$$

d'où, compte tenu de (4. 2) et de $\beta \circ \gamma = \alpha \circ \delta$,

$$\alpha \circ \delta({}^n c_{2i}) = \alpha({}^n C_{2i}), \qquad (1 \leq i \leq n),$$

et finalement $\delta({}^n c_{2i}) = {}^n C_{2i}$ puisque α est biunivoque,

Remarques. (1) Sans utiliser la formule de dualité on peut montrer aisément que ${}^nc_{2i} = \epsilon_i C_{2i}$, où $\epsilon_i = \pm 1$. Pour prouver la Proposition 4. 1, il ne reste donc plus qu'à établir les égalités $\epsilon_i = 1$ $(1 \leq i \leq n)$, ce qui peut se faire en se servant des valeurs des classes de Chern de la structure tangente de l'espace projectif complexe et du calcul des puissances de Steenrod des classes ${}^nC_{2i}$ qui sera fait au No. 11. Malheureusement cette méthode, au reste peu naturelle, conduit à des calculs assez compliqués.

Il serait intéressant de trouver une démonstration de la Proposition 4. 1 qui soit simple et indépendante de la dualité; peut être est-ce possible en utilisant l'expression des classes de Chern comme formes différentielles? On déduirait alors la formule de dualité directement de l'identité évidente (4. 2), comme cela est fait dans [3] pour les classes de Stiefel-Whitney réduites mod 2.

5. Autres groupes classiques. Nous allons passer brièvement en revue les autres groupes classiques, renvoyant à [2] pour plus de détails.[4]

Examinons tout d'abord les groupes *orthogonaux*. Le groupe $\boldsymbol{SO}(2n+1)$ est de rang n, et son groupe de Weyl est le groupe des permutations et changements de signes des x_i. Il en résulte donc que $H^*(B_{\boldsymbol{SO}(2n+1)}, Z_p)$ est l'algèbre des polynômes ayant comme générateurs les n éléments $P_{4i} = \Sigma x_1^2 \cdots x_i^2$, $(1 \leq i \leq n)$, lorsque p est impair (cette restriction étant due, rappelons-le, au fait que $\boldsymbol{SO}(n)$ a de la 2-torsion pour $m \geq 3$).

Le groupe $\boldsymbol{SO}(2n)$ est aussi de rang n, et son groupe de Weyl est engendré par les permutations et les changements de signes *en nombre pair* des x_i. Il en résulte aisément que pour p premier impair $H^*(B_{\boldsymbol{SO}(2n)}, Z_p)$ est l'algèbre des polynômes admettant comme générateurs les P_{4i} $(1 \leq i \leq n-1)$, et $W_{2n} = x_1 \cdots x_n$.

PROPOSITION 5. 1. *P_{4i} coïncide avec la classe de Pontrjagin de dimension $4i$, réduite mod p; W_{2n} coïncide avec la classe de Stiefel-Whitney de dimension $2n$, réduite mod p.*

Nous noterons p_{4i} et w_{2n} les classes de Pontrjagin et de Stiefel-Whitney respectivement.

Soit d'abord f l'inclusion de $U(n)$ dans $SO(2n)$; ces deux groupes étant de même rang, $\rho^*(f) = \rho^*(\boldsymbol{U}(n), \boldsymbol{SO}(2n))$ est biunivoque et, vu les identifications faites plus haut, se ramène au plongement des polynômes en les P_{4i}

[4] Pour tout ce qui concerne le groupe de Weyl, *voir* par exemple E. Stiefel, *Comm. Math. Helv.*, tome 14 (1941-42), pp. 350-380.

et W_{2n} dans l'algèbre de tous les polynômes symétriques. Il en résulte visiblement:

$$\rho^*(f)(W_{2n}) = C_{2n}. \tag{5.2}$$

Mais il est classique ([13], 41. 8) que, étant donné un espace fibré de groupe structural $\boldsymbol{U}(n)$ et de classes de Chern c_{2i}, on obtient en étendant le groupe structural à $\boldsymbol{SO}(2n)$ un espace fibré dont la classe de Stiefel-Whitney w_{2n} est égale à c_{2n}; cela signifie que $\rho^*(f)(w_{2n}) = c_{2n}$ et, comme $C_{2n} = c_{2n}$, l'égalité (5. 1) donne bien $W_{2n} = w_{2n}$, puisque $\rho^*(f)$ est biunivoque.

On pourrait raisonner de la même façon pour les classes de Pontrjagin, à l'aide du No. 3 de la Note [18] de Wu, mais il est plus commode de procéder différemment, en partant du plongement canonique de $\boldsymbol{SO}(n)$ dans $\boldsymbol{U}(n)$, qui définit un homomorphisme

$$\sigma = \rho^*(\boldsymbol{SO}(n), \boldsymbol{U}(n)) : H^*(B_{\boldsymbol{U}(n)}, Z_p) \to H^*(B_{\boldsymbol{SO}(n)}, Z_p)$$

déterminé explicitement dans [2], § 31; on a

$$\begin{aligned} &\sigma(C_{2i}) = 0 \text{ si } i \text{ est impair} \\ &\sigma(C_{4k}) = (-1)^k P_{4k}. \end{aligned} \tag{5.3}$$

D'autre part, on a d'après Wu ([20], p. 9):

$$\sigma(c_{4k}) = (-1)^k p_{4k},$$

ce qui, joint à $C_{4k} = c_{4k}$, donne $P_{4k} = p_{4k}$.

Remarque. De même que pour la Proposition 4. 1, il y aurait intérêt à avoir une démonstration de la Proposition 5. 1 indépendante des résultats de Wu, car on obtiendrait alors ces derniers de façon simple, par des calculs sur les polynômes symétriques et des changements de variables.

Le cas du groupe unitaire symplectique $\boldsymbol{Sp}(n)$ est tout à fait analogue à celui du groupe $\boldsymbol{SO}(2n+1)$, car ces deux groupes ont même rang et des groupes de Weyl isomorphes; la différence essentielle est que l'on peut raisonner directement avec des coefficients entiers puisque $\boldsymbol{Sp}(n)$ n'a pas de torsion.

On trouve alors que $H^*(B_{\boldsymbol{Sp}(n)}, Z)$ est identique à l'algèbre des polynômes ayant pour générateurs des classes K_{4i}, $(1 \leq i \leq n)$, définies par: $K_{4i} = \Sigma\, x_1^2 \cdots x_i^2$.

Le plongement canonique de $\boldsymbol{Sp}(n)$ dans $\boldsymbol{U}(2n)$ définit un homomorphisme

$$\nu = \rho^*(\boldsymbol{Sp}(n)), \boldsymbol{U}(2n)) \quad : \quad H^*(B_{\boldsymbol{U}(2n)}, Z) \to H^*(B_{\boldsymbol{Sp}(n)}, Z)$$

donné par les formules:

$$(5.4)\qquad \begin{aligned} \nu(C_{2i}) &= 0 \text{ si } i \text{ est impair} \\ \nu(C_{4k}) &= (-1)^k K_{4k}. \end{aligned}$$

Note. Les formules (5.2), (5.3), (5.4) permettent de déterminer les homomorphismes

$$H^*(\boldsymbol{SO}(2n), Z_p) \to H^*(\boldsymbol{U}(n), Z_p) \to H^*(\boldsymbol{SO}(n), Z_p)$$

et

$$H^*(\boldsymbol{U}(2n), Z) \to H^*(\boldsymbol{Sp}(n), Z)$$

induits par les plongements canoniques ([2], Proposition 21.3 et § 31). Ainsi, si l'on désigne par t_k l'élément de $H^{4k-1}(\boldsymbol{SO}(n), Z_p)$ dont l'image par transgression est P_{4k}, par u_n celui dont l'image est W_n (n pair), par v_k l'élément de $H^{4k-1}(\boldsymbol{Sp}(n), Z)$ dont l'image est K_{4k}, on voit que $(-1)^k t_k$ et $(-1)^k v_k$ sont restrictions des classes

$$h_{2k} \in H^{4k-1}(\boldsymbol{U}(n), Z_p), \text{ resp. } h_{2k} \in H^{4k-1}(\boldsymbol{U}(n), Z).$$

Par conséquent:

5.5. *A l'exception de la classe* u_n, (d'ailleurs exceptionnelle à plus d'un titre), *les générateurs des algèbres de cohomologie des groupes classiques sont restrictions de générateurs de l'algèbre de cohomologie du groupe unitaire (pour* $\boldsymbol{SO}(n)$, *on suppose* $p \neq 2$).

II. Les puissances réduites de Steenrod.

6. Les puissances réduites. Dans les Nos. 6 et 7, p désignera un nombre premier fixé, K un polyèdre fini, L un sous-polyèdre de K. On notera $H^q(K, L; Z_p)$ ou $H^q(K, L)$ lorsque cela ne prêtera pas à confusion, les groupes de cohomologie de K modulo L, à coefficients dans Z_p.

Les puissances réduites sont des homomorphismes

$$P_i^p\colon H^q(K, L; Z_p) \to H^{pq-i}(K, L; Z_p)$$

définis pour tout p premier, tout $i \geq 0$, tout $q \geq 0$, et tout couple de polyèdres K et L, L étant un sous-polyèdre de K.

A la place de la notation P_i^p, qui est adoptée dans [15], on trouve fréquemment ([4], [17], [20]) la notation St_p^i qui désigne P^p_{pq-q-i}; ainsi l'homomorphisme St_p^i applique $H^q(K, L)$ dans $H^{q+i}(K, L)$ et élève donc le degré de i unités.

Dans cet article, nous utiliserons une troisième notation, qui figure à la fin de [15b]. Rappelons les raisons de ce changement: D'après un théorème de Thom, ([16], [15b]), on a $St_p^i = 0$ si $i \not\equiv 0 \bmod 2\ (p-1)$ et il est connu que St_p^{2i+1} se ramène immédiatement à St_p^{2i}; ainsi, seules les opérations $St_p^{2k(p-1)}$ sont réellement importantes, et il est naturel de les désigner par un symbole plus simple; d'autre part les propriétés des St_p^i sont compliquées par la présence de facteurs numériques; si on cherche à les faire disparaître, en modifiant convenablement la définition des puissances réduites, on est finalement conduit à la notation suivante, qui est celle que nous adopterons dans toute la suite:

On désigne par $\mathcal{P}_p^k$ l'homomorphisme de $H^q(K, L; Z_p)$ dans

$$H^{q+2k(p-1)}(K, L; Z_p)$$

qui est égal à $\lambda(p, q, k) \cdot St_p^{2k(p-1)}$, le coefficient $\lambda(p, q, k)$ étant un élément de Z_p défini comme suit:

Si $p = 2,\ \lambda(p, q, k) = 1$

Si $p = 2h + 1,\ \lambda(p, q, k) = (-1)^{hr(r-1)/2}(h!)^{-r}$ *avec* $r = q - 2k$.

7. **Formulaire.** Nous rappelons ici les propriétés des puissances réduites $\mathcal{P}_p^k$, (*voir* [15b]); dans la suite de ce travail, nous n'utiliserons que ces formules, et jamais la définition explicite des $\mathcal{P}_p^k$, (ce qui n'est d'ailleurs pas surprenant, puisque Thom [17] a montré que les formules en question caractérisent complètement les $\mathcal{P}_p^k$).

7.1. $\mathcal{P}_p^k\colon H^q(K, L; Z_p) \to H^{q+2k(p-1)}(K, L; Z_p)$ *est un homomorphisme défini quels que soient* $q \geq 0$, $k \geq 0$ *et le couple* (K, L).

7.2. *Soit* $f\colon (K, L) \to (K', L')$ *une application continue. Si* f^* *est l'homomorphisme de* $H^*(K', L')$ *dans* $H^*(K, L)$ *induit par* f *on a* $\mathcal{P}_p^k \circ f^* = f^* \circ \mathcal{P}_p^k$.

7.3. *Lorsque* $p = 2$, *on a* $\mathcal{P}_p^k = Sq^{2k}$ *("i-carré" de Steenrod).*

7.4. $\mathcal{P}_p^0$ *est l'application identique de* $H^*(K, L)$ *sur lui-même.*

7.5. $\mathcal{P}_p^k\colon H^q(K, L) \to H^{q+2k(p-1)}(K, L)$ *est nul lorsque* $q < 2k$, *coïncide avec l'élévation à la* p*-ième puissance lorsque* $q = 2k$.

7.6. *Si* δ *désigne l'homomorphisme cobord qui applique* $H^q(L)$ *dans* $H^{q+1}(K, L)$, *on a* $\mathcal{P}_p^k \circ \delta = \delta \circ \mathcal{P}_p^k$.

7.7. *Soient x et y deux éléments de $H^*(K, L)$, $x \cdot y$ leur cup-produit. Lorsque $p \neq 2$, on a:*

$$\mathcal{P}_p^k(x \cdot y) = \Sigma_{i+j=k} \mathcal{P}_p^i(x) \cdot \mathcal{P}_p^j(y).$$

(Cela montre en particulier que $\mathcal{P}_p^1$ est une *dérivation.*)

7.8. $Sq^k(x \cdot y) = \Sigma_{i+j=k} \, Sq^i(x) \cdot Sq^j(y)$.

(*Ainsi, la formule 7.7 est valable pour $p = 2$ quand Sq^1 est nul pour tout élément de $H^*(K, L)$.*)

Remarquons enfin que la transgression dans un espace fibré, dont nous avons rappelé la définition au No. 2, est un produit $q^{*-1} \cdot \delta$, où q^* et δ commutent avec $\mathcal{P}_p^k$, vu 7.2 et 7.6; par conséquent:

7.9. *Si τ désigne la transgression dans un espace fibré E on a:*

$$\mathcal{P}_p^k \circ \tau = \tau \circ \mathcal{P}_p^k.$$

De façon plus précise, $\mathcal{P}_p^k$ applique $T^s(F)$ dans $T^{s+2k(p-1)}(F)$, et $L^{s+1}(B)$ dans $L^{s+1+2k(p-1)}(B)$, et on a un diagramme commutaif

$$\begin{array}{ccc} T^s(F) & \xrightarrow{\mathcal{P}_p^k} & T^{s+2k(p-1)}(F) \\ \downarrow \tau & & \downarrow \tau \\ H^{s+1}(B)/L^{s+1}(B) & \xrightarrow{\mathcal{P}_p^k} & H^{s+1+2k(p-1)}(B)/L^{s+1+2k(p-1)}(B) \end{array}$$

Note. Les puissances réduites $\mathcal{P}_p^k$ sont définies dans [15] pour les couples (K, L) de *polyèdres finis*; mais, comme nous l'a signalé N. E. Steenrod, elles sont définissables dans des théories cohomologiques plus vastes et notament dans la *théorie de Čech* (par passage à la limite à partir des polyèdres) et dans la *théorie singulière* (car il existe une formule simpliciale universelle, analogue à celle des i-produits, qui fait passer d'un cocycle représentant la classe de cohomologie x à un cocycle représentant $\mathcal{P}_p^k(x)$). Bien entendu, le formulaire précédent est encore valable dans ces deux cas.

8. Les puissances réduites dans les espaces projectifs. Nous allons montrer maintenant comment les formules du No. 7 permettent de déterminer les opérations $\mathcal{P}_p^k$ dans quelques cas simples. Les résultats obtenus nous serviront du reste dans la troisième partie de ce travail.

PROPOSITION 8.1. *Soit u une classe de cohomologie de dimension 2. On a, si $p \neq 2$, $\mathcal{P}_p^k(u^n) = \binom{n}{k} u^{n+k(p-1)}$ (on convient que $\binom{n}{k} = 0$ si $k > n$); si $p = 2$ la formule précédente reste valable lorsque $Sq^1 u = 0$.*

On raisonne par récurrence sur n. Supposons tout d'abord $p \neq 2$; en appliquant 7. 7 on obtient:

$$\mathcal{P}_p^k(u^n) = \Sigma_{i+j=k}\mathcal{P}_p^i(u) \cdot \mathcal{P}_p^j(u^{n-1})$$

et d'après l'hypothèse de récurrence, la somme du 2ème membre se réduit à

$$\binom{n-1}{k-1}u^{n+k(p-1)} + \binom{n-1}{k}u^{n+k(p-1)} = \binom{n}{k}u^{n+k(p-1)}.$$

Pour $p = 2$, on doit appliquer la formule 7. 8 et on trouve un terme supplémentaire, égal à $Sq^1u \cdot Sq^{2k-1}(u^{n-1})$, qui est nul si l'on suppose que $Sq^1u = 0$, et le reste du calcul vaut sans changement.

COROLLAIRE 8. 2. *Soient $\boldsymbol{X} = \boldsymbol{P}_m(C)$ l'espace projectif complexe de dimension complexe m, u un élément de $H^2(\boldsymbol{X}, Z_p)$. On a:*

$$\mathcal{P}_p^k(u^n) = \binom{n}{k}u^{n+k(p-1)}.$$

La seule chose à vérifier est que $Sq^1u = 0$, ce qui résulte de la nullité de $H^3(\boldsymbol{X}, Z_2)$.

COROLLAIRE 8. 3. *Soit $\boldsymbol{Y} = \boldsymbol{P}_m(K)$ l'espace projectif quaternionien de dimension quaternionienne m. Alors $H^4(\boldsymbol{Y}, Z_p)$ contient un élément $v \neq 0$ tel que:*

$$\mathcal{P}_p^k(v^n) = 0 \textit{ si } p = 2 \textit{ et si } k \textit{ est impair},$$

$$\mathcal{P}_p^k(v^n) = \binom{2n}{k}v^{n+k(p-1)/2} \textit{ sinon}.$$

Par définition même, $\boldsymbol{P}_m(K)$ est la base de $\boldsymbol{S}_{4m+3}$, fibrée par le groupe $\boldsymbol{Sp}(1)$ des quaternions de norme 1, qui est homéomorphe à $\boldsymbol{S}_3$; de même le quotient de $\boldsymbol{S}_{4m+3}$ par un sous-groupe $\boldsymbol{S}_1$ de $\boldsymbol{Sp}(1)$ est l'espace $\boldsymbol{P}_{2m+1}(C)$. Ce dernier est donc fibré de fibre $\boldsymbol{S}_3/\boldsymbol{S}_1 = \boldsymbol{S}_2$ et de base $\boldsymbol{P}_m(K)$; soit ψ la projection qui définit cette fibration. On voit immédiatement (soit par un raisonnement géométrique, soit en examinant la suite spectrale de cette fibration), que ψ^* est un isomorphisme de $H^*(\boldsymbol{Y})$ sur la sous-algèbre de $H^*(\boldsymbol{P}_{2m+1}(C))$ engendrée par u^2, u étant un générateur de $H^*(\boldsymbol{P}_{2m+1}(C))$. Soit alors v la réduction mod p de l'élément $\hat{v} \in H^4(\boldsymbol{Y}, Z)$ tel que $\psi^*(\hat{v}) = u^2$.

Pour $p = 2$, k impair, $\mathcal{P}_p^k(v^n)$ est de dimension congrue à 2 mod 4, et est donc nul; sinon on a

$$\psi^*(\mathcal{P}_p^k(v^n)) = \mathcal{P}_p^k(u^{2n}) = \binom{2n}{k}u^{2n+k(p-1)} = \phi^*(\binom{2n}{k}v^{n+k(p-1)/2}),$$

d'où le corollaire, puisque ψ^* est biunivoque.

Remarque. Il est naturel de se demander si, de même que dans le

Corollaire 8. 2, la formule du corollaire 8. 3 vaut pour *tout* élément de $H^4(\boldsymbol{Y})$. Un tel élément étant de la forme λv, $(\lambda \in Z_p)$, on est amené à voir si l'on a $\lambda^{k(p-1)/2} \equiv 1 \bmod p$; si k est pair, c'est bien le cas quel que soit $\lambda \not\equiv 0$, mais si k est impair, il faut et il suffit que λ soit reste quadratique mod p.

PROPOSITION 8. 4. *Soit X un espace dont l'algèbre de cohomologie $H^*(X, Z_p)$ est engendrée par des éléments de dimension 2. Si $(\mathcal{P}_p^1)^k$ désigne l'opération $\mathcal{P}_p^1$ itérée k fois, on a $(\mathcal{P}_p^1)^k(x) = k!\mathcal{P}_p^k(x)$ pour tout $x \in H^*(X, Z_p)$.*

Cette égalité résulte immédiatement de la Proposition 8. 1 pour x de dimension deux; il nous reste donc simplement à montrer que si elle est vraie pour x et y, elle l'est encore pour $x \cdot y$; or, $\mathcal{P}_p^1$ étant une dérivation d'après 7. 7, on peut lui appliquer la formule de Leibnitz donnant la dérivée k-ième d'un produit, et l'on obtient ainsi:

$$(\mathcal{P}_p^1)^k(x \cdot y) = \Sigma_{i+j=k} \tbinom{k}{i} (\mathcal{P}_p^1)^i(x) \cdot (\mathcal{P}_p^1)^j(y).$$

D'après l'hypothèse faite sur x et y, le second membre est égal à

$$\Sigma_{i+j=k}\, i!j! \tbinom{k}{i} \cdot \mathcal{P}_p^i(x) \cdot \mathcal{P}_p^j(y) = k!\, \Sigma_{i+j=k} \mathcal{P}_p^i(x) \cdot \mathcal{P}_p^j(y),$$

donc à $k!\mathcal{P}_p^k(x \cdot y)$, d'après 7. 7.

COROLLAIRE. *L'opération $\mathcal{P}_p^1$, itérée p fois, est nulle.*

Remarque. Si $p \neq 2$ la formule $(\mathcal{P}_p^1)^k = k!\mathcal{P}_p^k$ est très probablement valable sans hypothèse restrictive sur X; elle l'est en tout cas lorsque $X = G$ et $X = B_G$, G étant un groupe de Lie vérifiant les conditions 3. 2, car $H^*(B_G)$ est alors isomorphe à une sous-algèbre de $H^*(B_T)$, laquelle vérifie les hypothèses de la Proposition 8. 4, et on passe de là à $H^*(G)$ par transgression. Cette formule est par contre inexacte pour $p = 2$, comme le montre l'exemple $Sq^2 \circ Sq^2 = Sq^3 \circ Sq^1$.

9. Applications aux groupes d'homotopie des sphères. N. E. Steenrod, (*Reduced powers of cohomology classes,* Cours professé au Collège de France, Mai 1951), a montré comment on peut utiliser les puissances réduites pour étudier les groupes $\pi_i(\boldsymbol{S}_n)$. Rappelons sa méthode:

Soient p un nombre premier, k un entier, f une application continue de $\boldsymbol{S}_i$ dans $\boldsymbol{S}_n$, avec $i = n + 2k(p-1) - 1$. Désignons par $\boldsymbol{X}$ le complexe cellulaire obtenu en adjoignant à $\boldsymbol{S}_n$ une boule de dimension $i+1$ par l'application f de sa frontière dans $\boldsymbol{S}_n$; on a:

$$H^0(\boldsymbol{X}, Z) = H^n(\boldsymbol{X}, Z) = H^{n+2k(p-1)}(\boldsymbol{X}, Z) = Z$$

et les autres groupes de cohomologie de $\boldsymbol{X}$ sont nuls; soit encore s, (resp. t), la réduction mod p du générateur canonique de $H^n(\boldsymbol{X}, Z)$, (resp. de $H^{n+2k(p-1)}(\boldsymbol{X}, Z)$). On a $\mathcal{P}_p^k(s) = \lambda_f t$, $(\lambda_f \varepsilon Z_p)$, et il est clair que λ_f ne dépend que de la classe d'homotopie de f, et que l'application $f \to \lambda_f$ définit un homomorphisme de $\pi_{n+2k(p-1)-1}(\boldsymbol{S}_n)$ dans Z_p que nous noterons $\zeta_p^{n,k}$.

Si E désigne la suspension de Freudenthal, on a

9. 1 $$\zeta_p^{n+1,k} \circ E = \zeta_p^{n,k},$$

(cela résulte du fait que $\mathcal{P}_p^k$ commute avec δ).

Exemples.

1) $p = 2$. On déduit de l'existence d'applications dont l'invariant de Hopf est 1 que $\zeta_2^{n,1}$, $\zeta_2^{n,2}$ et $\zeta_2^{n,4}$ sont des homomorphismes de $\pi_{n+1}(\boldsymbol{S}_n)$, de $\pi_{n+3}(\boldsymbol{S}_n)$ et de $\pi_{n+7}(\boldsymbol{S}_n)$ *sur* Z_2 (pour $n \geq 2$, $n \geq 4$ et $n \geq 8$ respectivement).

2) $p = 3$. Il est classique que l'espace $\boldsymbol{X}$ obtenu par le procédé décrit plus haut à partir de l'application de Hopf $f : \boldsymbol{S}_7 \to \boldsymbol{S}_4$ est le *plan projectif quaternionien* $\boldsymbol{P}_2(K)$. Or, d'après 8. 3, on a $\mathcal{P}_3^1(v) \not\equiv 0 \bmod 3$ si v est un élément non nul de $H^4(\boldsymbol{P}_2(K), Z_3)$; par conséquent $\zeta_3^{4,1}(f) \neq 0$, ce qui, compte tenu de 9. 1, montre que $\zeta_3^{n,1}$ est un homomorphisme de $\pi_{n+3}(\boldsymbol{S}_n)$ *sur* Z_3 pour $n \geq 4$.

La suspension itérée E^2 étant un isomorphisme de $\pi_6(\boldsymbol{S}_3)$ sur un sous-groupe d'indice 2 de $\pi_8(\boldsymbol{S}_5)$, (d'après des résultats classiques de Freudenthal et de G. W. Whitehead), la formule 9. 1 montre que $\zeta_3^{3,1}$ applique $\pi_6(\boldsymbol{S}_3)$ *sur* Z_3, résultat obtenu d'une autre manière par Steenrod, et que nous préciserons au No. 19 en donnant explicitement un élément de $\pi_6(\boldsymbol{S}_3)$ dont l'image par $\zeta_3^{3,1}$ est $\neq 0$.

Ainsi, pour $p = 2, 3$, l'homomorphisme $\zeta_p^{n,1}$ applique $\pi_{n+2p-3}(\boldsymbol{S}_n)$ sur Z_p lorsque $n \geq 3$; nous allons voir que ce fait est général; plus précisément:

PROPOSITION 9. 2. *L'homomorphisme $\zeta_p^{n,1} : \pi_{n+2p-3}(\boldsymbol{S}_n) \to Z_p$ est un isomorphisme du p-composant de $\pi_{n+2p-3}(\boldsymbol{S}_n)$ sur Z_p lorsque $n \geq 3$.*

D'après [13], Chap. IV, Prop. 3 et 4, la suspension de Freudenthal applique isomorphiquement le p-composant de $\pi_{n+2p-3}(\boldsymbol{S}_n)$ sur celui de $\pi_{n+2p-2}(\boldsymbol{S}_{n+1})$ lorsque $n \geq 3$. Vu 9. 1, il suffit donc de démontrer 9. 2 pour $n = 3$; comme on sait que le p-composant de $\pi_{2p}(\boldsymbol{S}_3)$ est Z_p, (ibid. Proposition 7), on est finalement ramené à prouver que, si $f : \boldsymbol{S}_{2p} \to \boldsymbol{S}_3$ est une application essentielle définissant un élément d'ordre p de $\pi_{2p}(\boldsymbol{S}_3)$, on a $\lambda_f \neq 0$, avec les notations introduites au début de ce paragraphe.

Or, soit X l'espace obtenu en adjoignant à S_3 une cellule de dimension $2p+1$ à l'aide de f, et écrivons la suite exacte d'homotopie de (X, S_3) :

$$\cdots \to \pi_i(S_3) \to \pi_i(X) \to \pi_i(X, S_3) \xrightarrow{d} \pi_{i-1}(S_3) \to \pi_{i-1}(X) \to \cdots$$

On a visiblement $\pi_i(X, S_3) = 0$ lorsque $i < 2p+1$ et $\pi_{2p+1}(X, S_3) = Z$; en outre, l'image de $d : \pi_{2p+1}(X, S_3) \to \pi_{2p}(S_3)$ est le sous-groupe engendré par la classe de f, et est donc isomorphe à Z_p. Il suit de là:

$$\pi_i(X) \approx \pi_i(S_3) \text{ si } i < 2p, \qquad \pi_{2p}(X) \approx \pi_{2p}(S_3)/Z_p.$$

Soit alors $Y = (X, 4)$ l'espace obtenu à partir de X en tuant $\pi_3(X) = Z$, (au sens de [5], voir aussi [13], Chap. III). Par définition de Y, on a $\pi_i(Y) = 0$ pour $i \leq 3$, et $\pi_i(Y) = \pi_i(X)$ pour $i \geq 4$, ce qui, joint aux formules précédentes, montre que $\pi_i(Y)$, $(i \leq 2p)$, est un groupe fini dont le p-composant est nul; il en résulte que $H^i(Y, Z_p) = 0$ lorsque $0 < i \leq 2p$, ([13], Chap. III, Théor. 1).

Mais d'autre part Y est un espace fibré de base X et dont la fibre est un $K(Z, 2)$, au sens d'Eilenberg-MacLane. L'algèbre $H^*(Z, 2)$ est, comme on sait, une algèbre de polynômes engendrée par un élément de dimension deux, soit r. La transgression τ dans l'espace fibré Y transforme r en un élément de $H^3(X)$, qui est nécessairement de la forme λs, $(\lambda \in Z_p)$, s étant le générateur introduit plus haut. L'élément r étant transgressif, il en est de même de $r^r = \mathcal{P}_p^1(r)$, d'après 7. 9, et l'on a : $\tau(r^p) = \mathcal{P}_p^1(\tau(r)) = \lambda \mathcal{P}_p^1(s)$.

Si $\tau(r^p)$ étant nul, r^p définirait un élément non nul de $H^{2p}(Y, Z_p)$ ce qui est impossible, on l'a vu; on doit donc forcément avoir $\mathcal{P}_p^1(s) \neq 0$, ce qui signifie justement que $\lambda_f \neq 0$.

On remarquera que la démonstration précédente, à la différence de celles relatives à $p=2$ et à $p=3$, ne fournit aucun élément *explicite* de $\pi_{n+2p-3}(S_n)$ dont l'image par $\zeta_p^{n,1}$ soit non nulle.

III. Les puissances réduites dans la cohomologie des groupes de Lie et de leurs espaces classifiants.

10. Méthode générale. Nous revenons maintenant sur la méthode de calcul des puissances réduites dans $H^*(G, Z_p)$ et $H^*(B_G, Z_p)$ déjà brièvement décrite dans l'introduction.

p étant un nombre premier arbitraire, mais fixé, nous supposerons *dans toute la suite* que G vérifie les conditions 3. 2, autrement dit que G et son quotient G/T par un tore maximal sont sans p-torsion; comme nous l'avons

rappelé au No. 3, cela a lieu pour tout groupe classique et tout p, à l'exception du cas $G = \boldsymbol{SO}(n)$, $p = 2$; nous n'obtiendrons donc pas ici les Sq^i dans $H^*(\boldsymbol{SO}(n), Z_2)$ et $H^*(B_{\boldsymbol{SO}(n)}, Z_2)$, pour lesquels nous renvoyons à [3], où ils sont traités par une méthode analogue, utilisant des sous-groupes abéliens maximaux de type $(2, 2, \cdots, 2)$ au lieu de tores maximaux.

G vérifiant 3. 2, l'algèbre $H^*(B_G, Z_p)$ est une algèbre de polynômes à l générateurs de dimensions paires, soient $y_1, \cdots, y_l$, qui s'identifie canoniquement à une sous-algèbre de $H^*(B_T, Z_p)$; or cette dernière est une algèbre de polynômes à l générateurs $x_1, \cdots, x_l$ de dimension deux et les p-puissances réduites y sont déterminées par la Proposition 8. 1 et la formule 7. 7; cela résout donc la question pour $H^*(B_G, Z_p)$.

On passe de là à $H^*(G, Z_p)$ par transgression en utilisant les résultats du No. 2. Soit x_i l'élément universellement transgressif de $H^*(G, Z_p)$ tel que $\tau(x_i) = y_i \bmod D$, $(1 \leq i \leq l)$; d'après 7. 9, $\mathcal{P}_p^k(x_i)$ est aussi universellement transgressif et de plus, compte tenu de 2. 3:

$$\mathcal{P}_p^k(x_i) = \mathcal{P}_p^k(\tau x_i) = \mathcal{P}_p^k(y_i) \bmod D\tau;$$

mais les puissances réduites dans $H^*(B_G, Z_p)$ sont déjà connues; on sait donc exprimer $\mathcal{P}_p^k(y_i)$ comme polynôme en les y_j; désignons par $\Sigma \lambda_j y_j$ sa partie homogène de degré 1. On a donc $\mathcal{P}_p^k(y_i) = \Sigma \lambda_j y_j \bmod D$ ou encore

$$\mathcal{P}_p^k(y_i) = \Sigma \lambda_j(\tau x_j) = \tau(\Sigma \lambda_j x_j) \bmod D$$

d'où finalement $\mathcal{P}_p^k(x_i) = \Sigma \lambda_j x_j$, puisque τ est biunivoque. Cela détermine les puissances réduites des éléments universellement transgressifs de $H^*(G, Z_p)$; comme cette dernière est identique à l'algèbre extérieure engendrée par les x_i, les $\mathcal{P}_p^k$ s'y obtiennent alors grâce à 7. 7.

Remarque. On voit que la partie *essentielle* de cette méthode est le calcul des $\mathcal{P}_p^k$ dans $H^*(B_G, Z_p)$; pour en déduire $\mathcal{P}_p^k$ dans $H^*(G, Z_p)$, il nous suffit même de connaître le terme dominant de $\mathcal{P}_p^k(y_i)$, $(1 \leq i \leq l)$. Inversement la connaissance de $\mathcal{P}_p^k$ dans $H^*(G, Z_p)$ détermine le terme dominant de $\mathcal{P}_p^k(y_i)$ mais ne fournit aucun renseignement sur sa partie décomposable. En fait, nous n'aurons besoin que des termes dominants pour toutes les applications données dans la quatrième partie.

11. Le groupe unitaire $\boldsymbol{U(n)}$. Nous expliciterons tout d'abord la méthode générale dans le cas particulier le plus important, celui du groupe unitaire $\boldsymbol{U}(n)$. Nous reprenons les notations du No. 4; l'algèbre $H^*(B_{\boldsymbol{U}(n)}, Z_p)$ est engendrée par les classes C_{2i}, $(1 \leq i \leq n)$, et il s'agit essentiellement

d'exprimer $\mathcal{P}_p^k(C_{2i})$ comme polynôme en les C_{2i}; la classe C_{2i} s'identifie à la i-ième fonction symétrique élémentaire $\Sigma x_1 \cdots x_i = \sigma_i$, une fois $H^*(B_{U(n)}, Z_p)$ plongé dans $H^*(B_T, Z_p)$ comme il a été dit au No. 3.

LEMME 11. 1.

$$\mathcal{P}_p^k(x_1 \cdots x_i) = \Sigma_{1 \leq i_1 < \cdots < i_k \leq i}\, x^p_{i_1} \cdot x^p_{i_2} \cdot \cdots \cdot x^p_{i_k} \cdot x_{j_1} \cdot \cdots \cdot x_{j_{i-k}}$$

où $\{j_1 < j_2 < \cdots < j_{i-k}\}$ *est l'ensemble complémentaire de* $\{i_1, \cdots, i_k\}$ *dans la suite* $\{1, 2, \cdots, i\}$.

Démonstration par récurrence sur i; pour $i = 1$, le lemme se réduit à la Proposition 8. 1; d'après 7. 7, (qui vaut ici même si $p = 2$, car $H^*(B_T, Z_p)$ est nulle en toute dimension impaire), on a:

$$\mathcal{P}_p^k(x_1 \cdots x_i) = \mathcal{P}_p^k(x_1 \cdots x_{i-1}) \cdot x_i + \mathcal{P}_p^{k-1}(x_1 \cdots x_{i-1}) \cdot \mathcal{P}_p^1(x_i)$$

et, puisque $P_p^1(x_i) = x_i^p$ d'après 7. 5, cela donne:

$$\mathcal{P}_p^k(x_1 \cdots x_i) = \mathcal{P}_p^k(x_1 \cdots x_{i-1}) \cdot x_i + \mathcal{P}_p^{k-1}(x_1 \cdots x_{i-1}) \cdot x_i^p.$$

Vu l'hypothèse de récurrence, le premier terme du second membre est identique à la somme partielle de

$$\Sigma_{1 \leq i_1 < \cdots < i_k \leq i}\, x^p_{i_1} \cdot \cdots \cdot x_{i_k}^p \cdot x_{j_1} \cdot \cdots \cdot x_{j_{i-k}}$$

correspondant à $i_k < i$, et le second terme est identique à la somme partielle correspondant à $i_k = i$, ce qui démontre le lemme. (On observera que, si l'on y fait $x_1 = x_2 = \cdots = x_i = x$, on retrouve la Proposition 8. 1). Il résulte évidemment du lemme 11. 1 que

11. 2 $$\mathcal{P}_p^k(\Sigma x_1 \cdots x_i) = \Sigma x_1^p \cdots x_k^p x_{k+1} \cdots x_i,$$

d'où finalement

THÉORÈME 11. 3. *Soit* $B_p^{k,j}(\sigma_1, \cdots, \sigma_j)$, $(j = i + k(p-1))$, *le polynôme qui exprime le polynôme symétrique de terme typique* $x_1^p \cdots x_k^p x_{k+1} \cdots x_i$ *en fonction des* $\sigma_i = \Sigma x_1 \cdots x_i$. *Si* $C_{2i} \varepsilon H^{2i}(B_{U(n)}, Z_p)$ *désigne la classe de Chern de dimension* $2i$ *réduite mod* p, *on a*

$$\mathcal{P}_p^k(C_{2i}) = B_p^{k,j}(C_2, \cdots, C_{2j}).$$

Ce théorème est dû à Wu Wen Tsün [20]; notre démonstration est du reste tout à fait semblable à la sienne, la seule différence étant que chez Wu, l'égalité $C_{2i} = \Sigma x_1 \cdots x_i$ n'a qu'un caractère "symbolique" et ne peut être utilisée directement pour le calcul de $\mathcal{P}_p^k(C_{2i})$; (Wu raisonne par récurrence

sur n, en utilisant le théorème de dualité rappelé au No. 4, alors qu'ici nous avons interprété les x_i comme des éléments de $H^2(B_T, Z_p)$).

En combinant 10. 1 et 11. 3, on obtient:

COROLLAIRE 11. 4. *Soient $b_p^{k,j}\sigma_j$ le terme dominant de $B_p^k(\sigma_1, \cdots, \sigma_j)$, et $h_i \in H^{2i-1}(U(n), Z_p)$ l'élément dont l'image par transgression est C_{2i} mod D. On a:*

$$\mathcal{P}_p^k(h_i) = b_p^{k,j} h_j \qquad (1 \leq i \leq n;\, j = i + k(p-1)).$$

(Le terme dominant est bien entendu défini par la condition que

$$B_p^{k,j}(\sigma_1, \cdots, \sigma_j) - b_p^{k,j}\sigma_j$$

soit un polynôme en $\sigma_1, \cdots, \sigma_{j-1}$.)

12. Cas particuliers. Lorsque k, j, p sont des entiers donnés, le polynôme $B_p^{k,j}(\sigma_1, \cdots, \sigma_j)$ et son terme dominant $b_p^{k,j}\sigma_j$ peuvent être calculés par un procédé mécanique bien connu; pour $p = 2$, Wu Wen Tsün a même donné une formule générale, valable pour k et j quelconques:

$$B_2^{k,j} = \binom{j-k-1}{k}\sigma_j + \binom{j-k-2}{k-1}\sigma_1 \cdot \sigma_{j-1} + \cdots + \binom{j-2k}{1}\sigma_{k-1} \cdot \sigma_{j-k-1} + \sigma_k \cdot \sigma_{j-k};$$

nous ignorons s'il existe une formule générale du même genre pour $p \neq 2$.

Exemples.

1) Calcul de $B_3^{1,j}$. Il nous faut calculer $\Sigma\, x_1^3 x_2 \cdots x_{j-2}$; on a

$$\Sigma\, x_1^3 x_2 \cdots x_{j-2} = (\Sigma\, x_1^2) \cdot (\Sigma\, x_1 \cdots x_{j-2}) - \Sigma\, x_1^2 x_2 \cdots x_{j-1},$$

$$\Sigma\, x_1^2 x_2 \cdots x_{j-1} = (\Sigma\, x_1) \cdot (\Sigma\, x_1 \cdots x_{j-1}) - j \cdot \Sigma\, x_1 \cdots x_j.$$

Comme $\Sigma\, x_1^2 = (\sigma_1)^2 - 2\sigma_2$, on trouve en définitive:

12. 1 $$B_3^{1,j} = (\sigma_1)^2 \cdot \sigma_{j-2} - 2\sigma_2 \cdot \sigma_{j-2} - \sigma_1 \cdot \sigma_{j-1} + j \cdot \sigma_j,$$

d'où, compte tenu de 11. 3:

12. 2 $$\mathcal{P}_3^1(C_{2j-4}) = (C_2)^2 \cdot C_{2j-4} - 2 \cdot C_4 \cdot C_{2j-4} - C_2 \cdot C_{2j-2} + j \cdot C_{2j}.$$

2) Calcul de $b_p^{1,j}$. L'exemple précédent montre que $b_3^{1,j} = j$; nous allons voir que plus généralement:

12. 3 $$b_p^{1,j} \equiv j \mod p.$$

Puisque $(-1)^{p+1} \equiv 1 \bmod p$ pour tout p premier, il suffit évidemment

de prouver que le terme dominant de $\Sigma x_1^q x_2 \cdots x_{j-q+1}$ est $(-1)^{q+1} j \cdot \sigma_j$, ce qui, pour $q = 2$, résulte de la formule:

$$\Sigma x_1^2 x_2 \cdots x_{j-1} = (\Sigma x_1)(\Sigma x_1 \cdots x_{j-1}) - j \cdot \Sigma x_1 \cdots x_j,$$

et, pour $q > 2$, se démontre par récurrence sur q à l'aide de l'identité:

$$\Sigma x_1^q x_2 \cdots x_{j-q+1} = (\Sigma x_1^{q-1})(\Sigma x_1 \cdots x_{j-q+1}) - \Sigma x_1^{q-1} x_2 \cdots x_{j-q+2}.$$

En combinant 12. 3 avec 11. 3 et 11. 4, dont nous gardons les notations, on obtient:

PROPOSITION 12. 4. *Soient p un nombre premier, j un entier $> p$, non divisible par p. La classe de Chern C_{2j}, réduite mod p, est égale à $1/j \cdot \mathcal{P}_p^1(C_{2j-2p+2})$ augmentée d'un polynôme par rapport aux classes C_{2i}, $(i < j)$, et l'on a $h_j = 1/j \cdot \mathcal{P}_p^1(h_{j-p+1})$.*

En fait, dans la plupart des applications, c'est j qui est donné, et l'on cherche un nombre premier p vérifiant les hypothèses de 12. 4; on a à ce sujet:

LEMME 12. 5. *Pour tout entier $j \geq 3$, il existe un nombre premier p tel que $p < j$ et que $j \not\equiv 0 \bmod p$; ce nombre peut être choisi impair si $j \geq 4$.*

La première partie se démontre en prenant pour p un diviseur premier de $j-1$, la seconde en prenant pour p un diviseur premier de $j-1$ ou de $j-2$ suivant que j est pair ou impair.

Il résulte de 12. 4 et 12. 5:

PROPOSITION 12. 6. *Soit j un entier ≥ 3. Il existe un nombre premier $p < j$ ne divisant pas j, impair si $j \geq 4$, tel que C_{2j}, (resp. h_j), réduite mod p, soit égale à la somme d'un polynôme par rapport aux classes C_{2i}, (resp. h_i), $i < j$, et de $\mathcal{P}_p^1(\lambda C_{2j-2p+2})$, (resp. $\mathcal{P}_p^1(\lambda h_{j-p+1})$), $\lambda \in Z_p$.*

Puisque tout élément de $H^{2j}(B_{U(n)}, Z_p)$, (resp. de $H^{2j-1}(\boldsymbol{U}(n), Z_p)$, est somme d'un multiple de C_{2j}, (resp. de h_j), et d'un polynôme en les C_{2i}, (resp. en les h_i), $i < j$, on déduit de 12. 6:

COROLLAIRE 12. 7. *Si j et p vérifient les conditions de* 12. 6, *tout élément de $H^{2j}(B_{U(n)}, Z_p)$, et tout élément de $H^{2j-1}(\boldsymbol{U}(n), Z_p)$, peuvent s'exprimer à l'aide de cup-produits et d'opérations de Steenrod à partir d'éléments de dimensions strictement plus petites.*

On notera que ce Corollaire vaut aussi bien pour $\boldsymbol{SU}(n)$; cela pourrait se montrer par des calculs analogues, mais il est plus simple de remarquer que, $\boldsymbol{SU}(n)$ étant totalement nom homologue à zéro dans $\boldsymbol{U}(n)$, les algèbres

$H^*(\boldsymbol{SU}(n), Z_p)$ et $H^*(B_{SU(n)}, Z_p)$ sont des quotients de $H^*(\boldsymbol{U}(n), Z_p)$ et $H^*(B_{U(n)}, Z_p)$, ([2], Cor. à la Prop. 21. 3).

13. Le groupe symplectique unitaire $\boldsymbol{Sp(n)}$. Si l'on applique la méthode générale, on est amené à calculer $\mathcal{P}_p{}^k(K_{4i})$ avec $K_{4i} = \Sigma\, x_1{}^2 \cdots x_i{}^2$.

Le lemme 11. 1 donne la valeur de $\mathcal{P}_p{}^k(x_1{}^2 \cdots x_i{}^2)$, d'où celle de $\mathcal{P}_p{}^k(K_{4i})$; tous calculs faits, on trouve:

13. 1 $\mathcal{P}_p{}^k(\Sigma\, x_1{}^2 \cdots x_i{}^2)$

$$= \Sigma_{2r+s=k}\, 2^s \Sigma\, x_1{}^{2p} \cdot x_2{}^{2p} \cdots x_r{}^{2p} \cdot x_{r+1}{}^{p+1} \cdots x_{r+s}{}^{p+1} \cdot x_{r+s+1}{}^2 \cdots x_i{}^2.$$

Lorsque i et k sont assez petits, cette formule permet d'exprimer $\mathcal{P}_p{}^k(K_{4i})$ comme polynôme en les K_{4i}.

Exemples.

1) $k = 1$, $p = 3$. On doit calculer $2 \cdot \Sigma\, x_1{}^4 x_2{}^2 \cdots x_i{}^2$. On a

$$\Sigma\, x_1{}^4 x_2{}^2 \cdots x_i{}^2 = (\Sigma\, x_1{}^2)(\Sigma\, x_1{}^2 \cdots x_i{}^2) - (i+1)\Sigma\, x_1{}^2 \cdots x_{i+1}{}^2,$$

ce qui donne:

13. 2 $$\mathcal{P}_3{}^1(K_{4i}) = 2 \cdot K_4 \cdot K_{4i} - (2i+2) \cdot K_{4i+4}.$$

2) $k = 1$, $p = 5$. On doit calculer $2\,\Sigma\, x_1{}^6 x_2{}^2 \cdots x_i{}^2$. On a

$$\Sigma\, x_1{}^6 x_2{}^2 \cdots x_i{}^2 = (\Sigma\, x_1{}^4)(\Sigma\, x_1{}^2 \cdots x_i{}^2) - \Sigma\, x_1{}^4 x_2{}^2 \cdots x_{i+1}{}^2,$$

ce qui, compte tenu du calcul précédent et de $(\Sigma\, x_1{}^4) = (\Sigma\, x_1{}^2)^2 - 2 \cdot \Sigma x_1{}^2 x_2{}^2$ donne:

13. 3 $$\mathcal{P}_5{}^1(K_{4i}) = 2 \cdot K_4{}^2 \cdot K_{4i} + K_8 \cdot K_{4i} + 3 \cdot K_4 \cdot K_{4i+4} + (2i+4) \cdot K_{4i+8}.$$

En fait, il est en général plus commode d'utiliser le plongement canonique de $\boldsymbol{Sp}(n)$ dans $\boldsymbol{U}(2n)$; il conduit à un homomorphisme

$$\nu\colon H^*(B_{U(2n)}) \to H^*(B_{Sp(n)})$$

qui, d'après 5. 4, applique C_{4i+2} sur zéro et C_{4i} sur $(-1)^i K_{4i}$. On a donc:

$$\mathcal{P}_p{}^k(K_{4i}) = (-1)^i \mathcal{P}_p{}^k(\nu(C_{4i})) = (-1)^i \nu(\mathcal{P}_p{}^k(C_{4i})).$$

Appliquant alors le Théorème 11. 3, on trouve:

THÉORÈME 13. 4. *Avec les notations du Théorème* 11. 3, *on a:*

$$\mathcal{P}_p{}^k(K_{4i}) = (-1)^i B_p^{k,2j}(0, -K_4, 0, K_8, \cdots, 0, (-1)^t K_{4t}, \cdots, (-1)^j K_{4j}),$$

j désignant l'entier $i + k(p-1)/2$.

(On a supposé $p \neq 2$, ou bien k pair, car sinon il est évident que $\mathcal{P}_p^k(K_{4i}) = 0$, puisque $H^{4i+2k}(B_{Sp(n)}, Z_2) = 0$.)

De ce théorème on tire la conséquence suivante, analogue à 11. 4:

COROLLAIRE 13. 5. *Si v_i désigne l'élément de $H^{4i-1}(\boldsymbol{Sp}(n), Z_p)$ dont l'image par transgression est K_{4i} mod D, on a:*

$\mathcal{P}_p^k(v_i) = 0$ *si $p = 2$ et si k est impair,*

$\mathcal{P}_p^k(v_i) = (-1)^{k(p-1)/2} \cdot b_p^{k,2j} \cdot v_j$ *sinon, (avec $j = i + k(p-1)/2$).*

(On pourrait également déduire 13. 5 de 11. 4 et du fait que v_i est induit par $(-1)^i h_{2i}$, cf. No. 5).

Enfin, 13. 4 et 13. 5, joints à l'égalité $b_p^{1,2j} \equiv 2j \bmod p$, donnent l'analogue suivant de 12. 7:

COROLLAIRE 13. 6. *Soit j un entier > 2. Il existe un nombre premier $p < 2j$ impair tel que tout élément de $H^{4j}(B_{Sp(n)}, Z_p)$ et tout élément de $H^{4j-1}(\boldsymbol{Sp}(n), Z_p)$ puissent s'exprimer à l'aide de cup-produits et d'opérations de Steenrod à partir d'éléments de dimensions strictement plus petites.*

14. Le groupe orthogonal $\boldsymbol{SO}(n)$. Nous devons nous borner ici aux nombres premiers p *impairs*, puisque $\boldsymbol{SO}(n)$ a de la 2-torsion. On a vu au No. 5 que $H^*(\boldsymbol{SO}(n), Z_p)$ admet comme système de générateurs les classes de Pontrjagin réduites P_{4i}, auxquelles s'ajoute la classe de Stiefel-Whitney W_n lorsque n est pair; de plus le plongement canonique de $\boldsymbol{SO}(n)$ dans $\boldsymbol{U}(n)$ définit un homomorphisme $\sigma: H^*(B_{U(n)}, Z_p) \rightarrow H^*(B_{SO(n)}, Z_p)$ qui applique C_{4i+2} sur 0 et C_{4i} sur $(-1)^i P_{4i}$, (*voir* 5. 3). On peut alors répéter au sujet des P_{4i} le raisonnement fait au No. 13 pour les K_{4i}, et l'on obtient:

THÉORÈME 14. 1. *Si $P_{4i} \in H^{4i}(B_{SO(n)}, Z_p)$ désigne la classe de Pontrjagin de dimension $4i$ réduite mod p, $(p \neq 2)$, on a:*

$$\mathcal{P}_p^k(P_{4i}) = (-1)^i \cdot B_p^{k,2j}(0, -P_4, 0, P_8, \cdots, (-1)^j P_{4j}),$$
$$(j = i + k(p-1)/2).$$

Dans le cas où $n = 2m$ est pair, il reste encore à calculer $\mathcal{P}_p^k(W_{2m})$; le plus simple est ici d'appliquer la méthode générale; W_{2m} étant égale à $x_1 \cdots x_m$, on a $\mathcal{P}_p^k(W_{2m}) = \Sigma\, x_1^p \cdots x_k^p x_{k+1} \cdots x_m$, ou encore puisqu'il n'y a que m lettres $x_1 \cdots x_m$:

$$\mathcal{P}_p^k(W_{2m}) = x_1 \cdots x_m \Sigma\, x_1^{p-1} \cdots x_k^{p-1} = W_{2m} \Sigma\, x_1^{2h} \cdots x_k^{2h},$$

en posant $h = (p-1)/2$; il nous reste donc à exprimer $\Sigma\, x_1^{2h} \cdots x_k^{2h}$ comme

polynôme en les $P_{4i} = \Sigma_1{}^2 \cdots x_i{}^2$ et $(W_{2m})^2$, ce qui conduit au théorème suivant:

THÉORÈME 13. 2. *Soient p un nombre premier impair, $h = (p-1)/2$, et $C^{k,h}(\sigma_1, \cdots, \sigma_m)$ le polynôme qui exprime la fonction symétrique $\Sigma\, x_1{}^h \cdots x_k{}^h$ comme polynôme en les $\sigma_i = \Sigma\, x_1 \cdots x_i$, les lettres x_i étant au nombre de m. Si $W_{2m} \varepsilon H^{2m}(B_{SO(2m)}, Z_p)$ désigne la classe de Stiefel-Whitney de dimension $2m$, réduite mod p, on a:*

$$\mathcal{P}_p{}^k(W_{2m}) = W_{2m} \cdot C^{k,h}(P_4, P_8, \cdots, P_{4m-4}, (W_{2m})^2).$$

En particulier, on observera que $\mathcal{P}_p{}^k(W_{2m})$ est toujours un élément *décomposable* si $k \geq 1$.

Les théorèmes 13. 1 et 13. 2 entraînent:

COROLLAIRE 13. 3. *Soient p un nombre premier impair, t_i l'élément de $H^{4i-1}(\boldsymbol{SO}(n), Z_p)$ dont l'image par transgression est P_{4i} mod D, et, pour n pair, $u_n \varepsilon H^{n-1}(\boldsymbol{SO}(n), Z_p)$ celui dont l'image par transgression est W_n mod D. On a:*

$$\mathcal{P}_p{}^k(u_n) = 0 \text{ lorsque } k \geq 1.$$

$$\mathcal{P}_p{}^k(t_i) = (-1)^{k(p-1)/2} \cdot b_p{}^{k,2j} \cdot t_j, \qquad (j = i + k(p-1)/2).$$

(Remarquons en passant que l'on aurait pu prévoir *a priori* la nullité de $\mathcal{P}_p{}^k(u_n)$, $(k \geq 1)$. En effet, par définition même, W_n est image par transgression de l'élément de $H^{n-1}(\boldsymbol{SO}(n), Z)$ qui est image de la classe fondamentale de $\boldsymbol{S}_{n-1}$ par l'homomorphisme transposé de la projection naturelle de $\boldsymbol{SO}(n)$ sur $\boldsymbol{S}_{n-1}$; cet élément, une fois réduit mod p, est forcément égal à u_n puisque la transgression est ici biunivoque; $\mathcal{P}_p{}^k(u_n)$ est donc l'image d'un élément de $H^{n-1+2k(p-1)}(\boldsymbol{S}_{n-1}, Z_p)$ et est bien nul si $k \geq 1$.)

Enfin, par une démonstration tout à fait semblable à celle de 12. 7, on déduit de ce qui précède:

PROPOSITION 13. 4. *Soient j un entier ≥ 2, n un entier impair. Il existe un nombre premier impair $p < 2j$ tel que tout élément de $H^{4j}(B_{SO(n)}, Z_p)$ et tout élément de $H^{4j-1}(\boldsymbol{SO}(n), Z_p)$ puissent s'exprimer à l'aide de cup-produits et d'opérations de Steenrod à partir d'éléments de dimensions strictement plus petites.*

Remarque. Soit G un groupe classique, et soit $2q-1$ la plus grande dimension pour laquelle $H^*(G, R)$, (R corps des nombres réels), contient un élément universellement transgressif non nul.

Quel que soit le nombre premier impair $p < q$, l'opération $\mathcal{P}_p^1$ est non triviale dans $H^*(G, Z_p)$, et de façon plus précise, $\mathcal{P}_p^1$ transforme un élément $x \in H^3(G, Z_p)$ universellement transgressif et non nul, en un élément non nul de $H^{2p+1}(G, Z_p)$.

Cela se vérifie sur chaque cas particulier $G = \boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$, $\boldsymbol{Sp}(n)$, $\boldsymbol{SO}(n)$, en tenant compte de $b_p^{1,j} \equiv j \bmod p$; pour $G = \boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$ c'est encore vrai si $p = 2$.

Par contre, si $p > q$, il est évident a priori que $\mathcal{P}_p^1$, et, plus généralement, $\mathcal{P}_p^k$, est nul dans $H^*(G, Z_p)$.

IV. Applications.

15. Les sphères presque complexes et les algèbres à division sur le corps des nombres réels.

PROPOSITION 15. 1. *La sphère $\boldsymbol{S}_{2n}$, $(n \geq 4)$, n'admet pas de structure presque complexe.*

Raisonnons par l'absurde et soient $c_{2i} \in H^{2i}(\boldsymbol{S}_{2n}, Z)$ les classes de Chern d'une structure presque complexe de $\boldsymbol{S}_{2n}$; la classe c_{2i} est l'image de la classe $C_{2i} \in H^{2i}(B_{U(n)}, Z)$ par l'homomorphisme transposé d'une certaine application continue de $\boldsymbol{S}_{2n}$ dans $B_{U(n)}$, $(1 \leq i \leq n)$; il existe donc, d'après 12. 6 où l'on fait $j = n$, un nombre premier *impair* $p < n$ tel que c_{2n}, réduite mod p, s'exprime à l'aide de cup produits et de l'opération de Steenrod $\mathcal{P}_p^1$ à partir des c_{2i}, $(i < n)$. Mais ces dernières sont nulles puisque $H^{2i}(\boldsymbol{S}_{2n}, Z) = 0$ pour $0 < i < n$, ce qui montre que $c_{2n} \equiv 0 \bmod p$.

Soit d'autre part h la classe fondamentale de $H^{2n}(\boldsymbol{S}_{2n}, Z)$; on sait ([14], 41. 8), que c_{2n} est égale à $\chi(\boldsymbol{S}_{2n}) \cdot h$, c'est à dire à $2 \cdot h$. Comme on a choisi p impair, on voit que $c_{2n} \not\equiv 0 \bmod p$, d'où une contradiction.

Remarque. Le raisonnement précédent s'applique à la sphère $\boldsymbol{S}_{2n}$ munie *d'une structure différentiable quelconque*, alors que la démonstration classique d'inexistence d'une structure presque-complexe sur $\boldsymbol{S}_4$ ([14], 41. 20), suppose de façon essentielle que $\boldsymbol{S}_4$ est munie de la structure usuelle.

On sait que $\boldsymbol{S}_6$ admet une structure presque complexe définie à l'aide des octaves de Cayley ([14], 41. 21) ; en généralisant cette construction, nous allons démontrer la proposition suivante :

PROPOSITION 15. 2. *Soit A une algèbre (non necessairement associative) sur le corps des nombres réels R, jouissant des propriétés suivantes :*

(a) *A possède un élémente unité, qui sera noté e.*

(b) *La relation $a \cdot b = 0$ entraîne $a = 0$ ou $b = 0$.*

(c) *La relation $a \cdot b = e$ entraîne que a, b et e vérifient une relation linéaire à coefficients réels.*

Dans ces conditions, la dimension de l'algèbre A est 1, 2, 4, *ou* 8.

Avant de passer à la démonstration remarquons que, d'après Hopf et Stiefel,[5] les conditions (a) et (b) seules permettent d'établir que la dimension de A est une puissance de deux; on ignore si elles impliquent l'inégalité $\dim A \leq 8$. Dans le cas où la sous-algèbre engendrée par un élément quelconque est associative la condition (c) équivaut à dire que tout élément de A vérifie une relation quadratique.

Posons $n = \dim A$, et supposons $n \geq 3$; nous devons montrer que $n = 4$ ou $n = 8$. Introduisons sur A, considéré comme espace vectoriel réel, une forme quadratique définitive positive. Soient H l'hyperplan homogène de A orthogonal à e, (relativement au produit scalaire défini par cette forme), $\boldsymbol{S}$ l'ensemble des points de H à distance unité de l'origine; $\boldsymbol{S}$ est donc une sphère de dimension $n-2$. Si x est un point de $\boldsymbol{S}$ on peut identifier l'espace vectoriel T_x des vecteurs tangents à $\boldsymbol{S}$ en x au sous-espace de A orthogonal à e et x et de dimension $n-2$; soit k_x l'opération de projection orthogonale de A sur T_x, et posons

$$J_x(y) = k_x(x \cdot y) \quad \text{pour } y \in T_x.$$

L'opérateur J_x est un endomorphisme de T_x qui varie continûment avec x; montrons que *J_x n'a pas de valeur propre réelle*: une égalité $J_x(y) = \lambda y$, $(\lambda \in R)$, peut s'écrire $k_x(x \cdot y - \lambda e) = 0$, ou encore $x \cdot y - \lambda y = \mu e + \nu x$, $(\mu, \nu \in R)$, ce qui donne $(x - \lambda \cdot e)(y - \nu \cdot e) = (\mu + \lambda \cdot \nu) e$. Si l'on suppose $y \neq 0$, les éléments $x - \lambda \cdot e$ et $y - \nu \cdot e$ sont $\neq 0$ puisque e et y sont orthogonaux à e; il s'ensuit d'après (b) que $(\mu + \lambda \cdot \nu) \neq 0$; on peut donc poser $w = (\mu + \lambda \cdot \nu)^{-1} \cdot (y - \nu e)$ et l'on a $(x - \lambda \cdot e) \cdot w = e$; d'après (c), il existe alors une relation linéaire entre $(x - \lambda \cdot e)$, w et e, donc aussi entre x, y et e; mais cela est absurde puisque ces éléments sont deux à deux orthogonaux.

Ainsi, nous avons associé à tout point $x \in \boldsymbol{S}$ un endomorphisme sans valeurs propres réelles J_x de l'espace des vecteurs tangents à $\boldsymbol{S}$ en x; il en résulte classiquement (voir ci-dessous) que $\boldsymbol{S}$ peut être munie d'une structure presque complexe, et l'on a donc $n - 2 = 2$ ou 6, c'est à dire $n = 4$ ou 8.

Note. Indiquons encore, pour être complet, comment on passe de l'exis-

[5] E. Stiefel, *Comm. Math. Helv.*, tome 13 (1940-41), pp. 201-218, H. Hopf, *ibid.*, pp. 219-239.

tence de l'endomorphisme J_x à une structure presque complexe sur **S**.[6] Il nous faut remplacer l'endomorphisme J_x par un endomorphisme I_x tel que $(I_x)^2 = -1$.

Soit $T_x \otimes C$ l'extension complexe de l'espace vectoriel réel T_x et soient $(\alpha_1, \cdots, \alpha_q, \bar{\alpha}_1, \cdots, \bar{\alpha}_q)$ les valeurs propres de J_x dans $T_x \otimes C$, chaque α_i ayant une partie imaginaire positive; pour toute valeur propre α nous désignons par V_α le plus grand sous-espace vectoriel de $T_x \otimes C$ sur lequel $J_x - \alpha$ est nilpotent; on sait que $T_x \otimes C$ est la somme directe des V_α, et il est clair que $V_{\bar{\alpha}} = \bar{V}_\alpha$; l'espace $T_x \otimes C$ est donc la somme directe de $W = V_{\alpha_1} + \cdots + V_{\alpha_q}$ et de $\bar{W}$, et tout élément $y \in T_x$ peut se mettre d'une seule façon sous la forme $y = w + \bar{w}$, $(w \in W)$. Posons alors $I_x(y) = iw - i\bar{w}$, c'est un élément de T_x et ainsi I_x définit un endomorphisme de T_x qui vérifie visiblement la condition $(I_x)^2 = -1$; il reste encore à s'assurer qu'il varie continûment avec x; cela résulte, par exemple, de la continuité des valeurs propres de J_x.

16. Sur les espaces fibrés à base sphérique. Nous intercalons ici quelques résultats dont nous aurons besoin dans les Nos. suivants; la Proposition 16.1 a été également utilisée par Miller [10].

PROPOSITION 16. 1. *Soient E un espace fibré de fibre F, de base S_r, β la classe fondamentale de $H^r(S_r, Z_p)$, (p premier), ζ la projection de E sur S_r. Si la classe $\gamma = \zeta^*(\beta)$ peut s'exprimer à l'aide de cup-produits et d'opérations de Steenrod à partir d'éléments de $H^*(E, Z_p)$ de dimensions $< r$, l'espace fibré E n'admet pas de section.*

Par hypothèse on peut trouver des éléments $u_1, \cdots, u_k \in H^*(E, Z_p)$, $(0 < \dim u_i < r)$, tels que $\gamma = f(u_1, \cdots, u_k)$, où f est une expression formée à l'aide de cup-produits et d'operations $\mathcal{P}_p^k$. Si $s: S_r \to E$ était une section de l'espace fibré E, l'homomorphisme $s^*: H^*(E, Z_p) \to H^*(S_r, Z_p)$ vérifierait la relation $s^* \circ \zeta^* = 1$, et l'on aurait:

$$\beta = s^* \circ \zeta^*(\beta) = s^*(\gamma) = s^*(f(u_1, \cdots, u_k)) = f(s^*(u_1), \cdots, s^*(u_k)) = 0$$

puisque $s^*(u_i)$ est une classe de cohomologie de S_r de dimension strictement comprise entre 0 et r; mais comme $\beta \neq 0$, cela est impossible et montre qu'il n'y a pas de section.

(Bien entendu, ce genre de raisonnement a une portée plus générale et s'applique à d'autres espaces que les sphères, pourvu que l'on soit certain que $s^*(u_i) = 0$ pour tout i.)

La Proposition 16. 1 revient à dire que *la classe caractéristique* $\alpha \in \pi_{r-1}(F)$

[6] La démonstration qui suit nous a été obligeamment communiquée par G. de Rham.

de la fibration considérée est un élément non nul de $\pi_{r-1}(F)$; nous allons préciser ce résultat dans les deux propositions suivantes.

PROPOSITION 16.2. *Avec les hypothèses et notations précédentes, la classe caractéristique* $\alpha \varepsilon \pi_{r-1}(F)$ *est, soit d'ordre infini, soit d'ordre fini divisible par* p.

Il nous faut montrer que l'on a $q \cdot \alpha \neq 0$ dans $\pi_{r-1}(F)$ lorsque q est un entier $\neq 0$ et non divisible par p.

Pour cela, soient $\psi : S_r \to S_r$ une application de degré q, E' l'espace fibré image réciproque de E par ψ, et ζ' la projection de E' sur S_r; il existe donc une application $\bar{\psi} : E' \to E$ telle que $\psi \circ \zeta' = \zeta \circ \bar{\psi}$. On a évidemment $\psi^*(\beta) = q \cdot \beta$, d'où

$$q \cdot \zeta'^*(\beta) = \zeta'^* \circ \psi^*(\beta) = \bar{\psi}^* \circ \zeta^*(\beta) = \bar{\psi}^*(f(u_1, \cdots, u_k)),$$

ce que donne, puisque $q \not\equiv 0 \bmod p$,

$$\zeta'^*(\beta) = 1/q \cdot f(\bar{\psi}^*(u_1), \cdots, \bar{\psi}^*(u_k)),$$

et E' n'a pas de section d'après 16. 1 ; cela signifie que la classe caractéristique $\alpha' \varepsilon \pi_{r-1}(F)$ de E' est non nulle, et comme cette classe est évidemment égale à $q \cdot \alpha$, la proposition est démontrée.

PROPOSITION 16.3. *Ajoutons aux hypothèses de* 16. 1 *les suivantes:*

(a) *L'espace fibré* E *est un espace fibré principal à groupe structural* F *connexe par arcs.*

(b) *Si* $u_1, \cdots, u_k$ *sont les éléments de* $H^*(E, Z_p)$ *tels que* $\gamma = f(u_1, \cdots, u_k)$, *on a* $0 < \dim u_i \leq r - 2$ *pour tout* i.

(c) *La classe* γ *est* $\neq 0$.

Alors il n'existe aucun élément $\alpha' \varepsilon \pi_{r-1}(F)$ *tel que* $\alpha = p \cdot \alpha'$.

(Autrement dit, α définit un élément non nul de $\pi_{r-1}(F) \otimes Z_p$, résultat évidemment plus précis que celui de 16. 2.)

Nous raisonnons par l'absurde et supposons donc l'existence d'un élément $\alpha' \varepsilon \pi_{r-1}(F)$ tel que $\alpha = p \cdot \alpha'$. On sait ([14], 18. 5), que si l'on associe à tout espace fibré principal de base S_r, de fibre F, sa classe caractéristique, on définit une correspondance biunivoque entre les classes d'espaces fibrés principaux de fibre F, base S_r, et les éléments de $\pi_{r-1}(F)$. Vu notre hypothèse, il existe donc un espace fibré principal E', de classe caractéristique α', dont E est l'image réciproque par une application $\sigma : S_r \to S_r$ de degré p. On désigne comme dans la démonstration précédente par ζ' la projection de E' sur S_r; soient encore $\bar{\sigma} : E \to E'$ l'application canonique de E dans E'

et $i^*: H^*(E, Z_p) \to H^*(F, Z_p)$, resp. $i'^*: H^*(E', Z_p) \to H^*(F, Z_p)$, l'homomorphisme défini par l'injection d'une fibre dans E, resp. E'.

On a évidemment $i^* \circ \bar{\sigma}^* = i'^*$; mais, d'après la suite exacte de H. C. Wang (*Duke Math. Jour.*, vol. 16 (1949), 33-38, ou [12], p. 471), i^* et i'^* sont des isomorphismes sur pour les dimensions $\leq r - 2$; il en est donc de même pour $\bar{\sigma}^*$ et vu l'hypothèse (b), $H^*(E', Z_p)$ contient des éléments u'_i tels que $u_i = \bar{\sigma}^*(u'_i)$, $(1 \leq i \leq k)$. On a donc

$$i^*(\gamma) = i^*(f(u_1, \cdots, u_k)) = i^* \circ \bar{\sigma}^*(f(u'_1, \cdots, u'_k)) = i'^*(f(u'_1, \cdots, u'_k)).$$

Il est clair que $i^*(\gamma) = 0$, donc $i'^*(f(u'_1, \cdots, u'_k)) = 0$, et la suite de Wang donne alors $f(u'_1, \cdots, u'_k) = \lambda \cdot \zeta'^*(\beta)$, $(\lambda \in Z_p)$. On en tire

$$\gamma = \lambda \cdot \bar{\sigma}^* \circ \zeta'^*(\beta) = \lambda \cdot \zeta^* \circ \sigma^*(\beta),$$

mais cela est impossible car $\sigma^*(\beta) = 0$ puisque σ est de degré p, et d'autre part $\gamma \neq 0$ d'après (c), ce qui démontre 16. 3.

17. Inexistence de sections dans certains espaces fibrés.

PROPOSITION 17. 1. *Les fibrations suivantes n'ont pas de section:*

(a) $\boldsymbol{SU}(n)/\boldsymbol{SU}(n-1) = \boldsymbol{S}_{2n-1}$ *pour* $n \geq 3$.

(b) $\boldsymbol{U}(n)/\boldsymbol{U}(n-1) = \boldsymbol{S}_{2n-1}$ *pour* $n \geq 3$.

(c) $\boldsymbol{Sp}(n)/\boldsymbol{Sp}(n-1) = \boldsymbol{S}_{4n-1}$ *pour* $n \geq 2$.

(d) $\boldsymbol{Spin}(9)/\boldsymbol{Spin}(7) = \boldsymbol{S}_{15}$.

(e) $\boldsymbol{Spin}(7)/\boldsymbol{G}_2 = \boldsymbol{S}_7$.

(Les fibrations (a), (b), (c) sont classiques, pour les deux dernières *voir* [1].)

D'après 16. 1, il suffit de trouver dans chaque cas un nombre premier p tel que tout élément de $H^{2n-1}(\boldsymbol{SU}(n), Z_p)$, de $H^{2n-1}(\boldsymbol{U}(n), Z_p)$, de $H^{4n-1}(\boldsymbol{Sp}(n), Z_p)$, de $H^{15}(\boldsymbol{Spin}(9), Z_p)$, de $H^7(\boldsymbol{Spin}(7), Z_p)$ s'exprime à l'aide de cup-produits et de puissances réduites à partir d'éléments de dimensions strictement inférieures. C'est possible pour $\boldsymbol{U}(n)$ et $\boldsymbol{SU}(n)$ d'après 12. 7 et pour $\boldsymbol{Sp}(n)$ d'après 13. 6. Dans le cas (d), on choisit d'abord p impair tel que tout élément de $H^{15}(\boldsymbol{SO}(9), Z_p)$ s'exprime à l'aide de cup-produits et d'opérations $\mathcal{P}_p^k$ à partir d'éléments de dimensions < 15, ce qui est possible pour $p = 3, 5, 7$ d'après 13.4; puisque $\boldsymbol{Spin}(9)$ est un revêtement à deux feuillets de $\boldsymbol{SO}(9)$, l'homomorphisme $H^*(\boldsymbol{SO}(9), Z_p) \to H^*(\boldsymbol{Spin}(9), Z_p)$ transposé de la projection est un isomorphisme sur, et la même propriété a lieu dans $H^*(\boldsymbol{Spin}(9), Z_p)$; on raisonne de la même façon dans le cas (e), pour $p = 3$.

Remarque. La Prop. 17. 1 (a) montre que $\boldsymbol{SU}(3)/\boldsymbol{SU}(2)=\boldsymbol{S}_5$ n'a pas de section, autrement dit que la classe caractéristique $\alpha\,\varepsilon\,\pi_4(\boldsymbol{S}_3)$ de cette fibration est non nulle, ce qui a été tout d'abord démontré par Pontrjagin [11], par une étude homotopique de α; on observera que nous sommes parvenus à ce résultat par voie cohomologique, (en utilisant l'opération Sq^2).

PROPOSITION 17. 2. (a) *La classe caractéristique de la fibration* $\boldsymbol{SU}(n)/\boldsymbol{SU}(n-1)=\boldsymbol{S}_{2n-1}$ *définit un élément non nul de* $\pi_{2n-2}(SU(n-1))\otimes Z_p$ *pour tout p premier* $<n$, *ne divisant pas n.*

(b) *Il en est de même pour la fibration* $\boldsymbol{U}(n)/\boldsymbol{U}(n-1)=\boldsymbol{S}_{2n-1}$.

(c) *Il en est de même pour tout p premier impair* $<2n$, *ne divisant pas n, pour la fibration* $\boldsymbol{Sp}(n)/\boldsymbol{Sp}(n-1)=\boldsymbol{S}_{4n-1}$.

(d) *La classe caractéristique de la fibration* $\boldsymbol{Spin}(9)/\boldsymbol{Spin}(7)=\boldsymbol{S}_{15}$ *définit un élément non nul de* $\pi_{14}(Spin(7))\otimes Z_p$ *pour* $p=3,5,7$.

(e) *La classe caractéristique de la fibration* $\boldsymbol{Spin}(7)/\boldsymbol{G}_2=\boldsymbol{S}_7$ *définit un élément non nul de* $\pi_6(G_2)\otimes Z_3$.

On doit montrer que les hypothèses de 16. 3 sont vérifiées; 16. 3(a) est évidente, 16. 3(b) résulte de ce que l'opération f utilisée pour établir 17. 1 est chaque fois $\mathcal{P}_p{}^1$, qui augmente le degré $2(p-1)\geq 2$ unités. Enfin, il est évident dans chaque cas envisagé ici que γ est universellement transgressif, et il résulte de la détermination même des algèbres $H^*(E,Z_p)$ que $\gamma\neq 0$; 16. 3(c) est donc aussi satisfaite.

On vérifie ensuite, en utilisant la formule $b_p{}^{1,j}\equiv j \bmod p$ et les résultats de III que dans chaque cas, les nombres premiers de l'énoncé sont tels que $\gamma=\mathcal{P}_p{}^1(u)$, $(u\,\varepsilon\,H^*(E,Z_p))$; 17. 2 résulte donc de 16. 3.

Remarque. Pour démontrer la Proposition 17. 2, nous n'avons eu besoin que des opération $\mathcal{P}_p{}^1$. En se servant des puissances réduites $\mathcal{P}_p{}^k$ pour k quelconque, et d'un lemme sur les coefficients $b_p{}^{k,j}$ qui sera établi plus loin (lemme 20. 7), on déduit par le raisonnement ci-dessus un résultat plus complet, que nous énoncerons uniquement dans le cas (a) pour simplifier:

La classe caractéristique de $\boldsymbol{SU}(n)/\boldsymbol{SU}(n-1)=\boldsymbol{S}_{2n-1}$ *définit un élément non nul de* $\pi_{2n-2}(\boldsymbol{SU}(n-1))\otimes Z_p$ *lorsque le nombre premier* $p<n$ *vérifie la condition suivante: Si* $h(p,n)$ *est le plus grand entier tel que* $p^{h(p,n)}\cdot(p-1)<n$, *le nombre n n'est pas divisible par* $p^{h(p,n)+1}$.

Nous terminons le No. 17 par la Proposition suivante, tout à fait analogue au cas (c) des Prop. 17. 1 et 17. 2:

PROPOSITION 17. 3. *Soit* $\boldsymbol{W}_{4n-1}$ *la variété des vecteurs de longueur unité tangents à la sphère* $\boldsymbol{S}_{2n}$. *La fibration* $\boldsymbol{SO}(2n+1)/\boldsymbol{SO}(2n-1)=\boldsymbol{W}_{4n-1}$

n'a pas de section si $n \geq 2$. *Si* p *est un nombre premier impair* $< 2n$, *ne divisant pas* n, *l'homomorphisme bord* d *de* $\pi_{4n-1}(\boldsymbol{W}_{4n-1})$ *dans* $\pi_{4n-2}(\boldsymbol{SO}(2n-1))$ *définit un sous-groupe isomorphe à* Z_p *de* $\pi_{4n-2}(\boldsymbol{SO}(2n-1)) \otimes Z_p$.

On sait que, si p est impair, $H^*(\boldsymbol{W}_{4n-1}, Z_p) \approx H^*(\boldsymbol{S}_{4n-1}, Z_p)$; le raisonnement de la Prop. 16. 1 s'applique donc sans changement et la première partie de la Proposition résulte de 13. 4.

On sait (*voir* [13], Chap. IV, Prop. 2), qu'il existe une application g de $\boldsymbol{S}_{4n-1}$ dans $\boldsymbol{W}_{4n-1}$ telle que le noyau et le conoyau des applications

$$g_0 : \pi_i(\boldsymbol{S}_{4n-1}) \rightarrow \pi_i(\boldsymbol{W}_{4n-1})$$

soient des 2-groupes pour toute valeur de i. Soit $\boldsymbol{E}$ l'espace fibré image réciproque de $\boldsymbol{SO}(2n+1)$ par cette application, et $\bar{g}$ son application canonique dans $\boldsymbol{SO}(2n+1)$. Comme la restriction de $\bar{g}$ à une fibre est un homéomorphisme sur une fibre de $\boldsymbol{SO}(2n+1)$, et comme g^* est pour tout $p \neq 2$ un isomorphisme de $H^*(\boldsymbol{S}_{4n-1}, Z_p)$ sur $H^*(\boldsymbol{W}_{4n-1}, Z_p)$, l'homomorphisme $\bar{g}^*$ est un isomorphisme de $H^*(\boldsymbol{SO}(2n+1), Z_p)$ sur $H^*(\boldsymbol{E}, Z_p)$, ([2], § 4d) ; on peut donc appliquer à $\boldsymbol{E}$ les mêmes raisonnements et calculs qu'à la fibration (c) de 17. 2. Par conséquent l'image de

$$d' : \pi_{4n-1}(\boldsymbol{S}_{4n-1}) \rightarrow \pi_{4n-2}(\boldsymbol{SO}(2n-1)) \otimes Z_p$$

est un sous-groupe de $\pi_{4n-2}(\boldsymbol{SO}(2n-1)) \otimes Z_p$ isomorphe à Z_p, ce qui, joint à l'égalité $d' = d \circ g_0$, démontre la deuxième partie de 17. 3.

18. Sur les *p*-composants des groupes d'homotopie des groupes classiques. Les Propositions précédentes permettent de calculer les p-composants des groupes $\pi_i(G)$, $(G = \boldsymbol{SU}(n), \boldsymbol{Sp}(n), \boldsymbol{SO}(n))$, jusqu'à la dimension $4p-3$. Pour énoncer les résultats obtenus, il sera commode d'utiliser le langage de la C-théorie de [13], et en particulier la notion de C-isomorphisme définie dans le Chap. I de [13]. Nous désignons par C_p la classe des groupes finis d'ordre premier à p.

PROPOSITION 18. 1. *Soient* p *un nombre premier,* $\boldsymbol{G} = \boldsymbol{SU}(n)$. *A un* C_p*-isomorphisme près, les groupes* $\pi_i(\boldsymbol{G})$, $(i \leq 4p-3)$, *sont les suivants:*

(I). *Si* $n \leq p$, *on a* $\pi_{2j-1}(\boldsymbol{G}) \equiv Z$, $(2 \leq j \leq n)$, $\pi_{2k}(\boldsymbol{G}) \equiv Z_p$, $(p \leq k \leq p+n-2)$, $\pi_{4p-3}(\boldsymbol{G}) \equiv Z_p$, $\pi_i(\boldsymbol{G}) \equiv 0$ *sinon.*

(II). *Si* $p < n \leq 2p-2$, *on a* $\pi_{2j-1}(\boldsymbol{G}) \equiv Z$, $(2 \leq j \leq n)$. $\pi_{2k}(\boldsymbol{G}) \equiv Z_p$, $(n \leq k \leq 2p-2)$, $\pi_i(\boldsymbol{G}) \equiv 0$ *sinon.*

(III). *Si* $n \geq 2p-1$, *on a* $\pi_{2j-1}(\boldsymbol{G}) \equiv Z$, $(2 \leq j \leq 2p-1)$, $\pi_i(\boldsymbol{G}) \equiv 0$ *sinon.*

Si $n \leq p$, il résulte de [13], Chap. V, Prop. 6 que p est *régulier* pour $\boldsymbol{G}$, (au sens du § 4 de cet article), et il s'ensuit que $\pi_i(\boldsymbol{G})$ est C_p-isomorphe à la somme directe des groupes $\pi_i(\boldsymbol{S}_{2m-1})$, $(2 \leq m \leq n)$, ce qui démontre (I).

A partir de là, on raisonne par récurrence sur n pour établir (II) et (III). Il suffit pour cela de considérer la suite exacte:

$$\pi_{i+1}(\boldsymbol{S}_{2n-1}) \to \pi_i(\boldsymbol{SU}(n-1)) \to \pi_i(\boldsymbol{SU}(n)) \to \pi_i(\boldsymbol{S}_{2n-1}),$$

en remarquant que les groupes $\pi_i(\boldsymbol{S}_{2n-1})$ sont tous C_p-nuls sauf pour $i = 2n-1$ puisque $2n-1+2p-3 \geq 4p-2$ et en utilisant la Proposition 17. 2 pour déterminer l'image de $d: \pi_{2n-1}(\boldsymbol{S}_{2n-1}) \to \pi_{2n-2}(\boldsymbol{SU}(n-1))$ quand $p < n \leq 2p-2$.

Le lecteur n'aura pas de peine à obtenir des résultats analogues pour $G = \boldsymbol{Sp}(n), \boldsymbol{SO}(2n+1)$ que nous n'expliciterons pas. Nous nous bornerons à indiquer comment l'on passe de $\boldsymbol{SO}(2n-1)$ à $\boldsymbol{SO}(2n)$:

PROPOSITION 18. 2. *Si C désigne la classe des 2-groupes, le groupe $\pi_i(\boldsymbol{SO}(2n))$ est C-isomorphe à la somme directe de $\pi_i(\boldsymbol{SO}(2n-1))$ et de $\pi_i(\boldsymbol{S}_{2n-1})$.*

On sait que la classe caractéristique α de la fibration $\boldsymbol{SO}(2n)/\boldsymbol{SO}(2n-1) = \boldsymbol{S}_{2n-1}$ vérifie $2 \cdot \alpha = 0$. L'espace fibré $\boldsymbol{E}$, image réciproques de $\boldsymbol{SO}(2n)$ par une application de $\boldsymbol{S}_{2n-1}$ sur $\boldsymbol{S}_{2n-1}$ de degré deux, est donc isomorphe à $\boldsymbol{SO}(2n-1) \times \boldsymbol{S}_{2n-1}$. En outre, on tire de [2], § 4d, exactement comme dans la démonstration de 17. 3, que l'application canonique de $\boldsymbol{E}$ sur $\boldsymbol{SO}(2n)$ définit un isomorphisme de $H^*(\boldsymbol{SO}(2n), Z_p)$ sur $H^*(\boldsymbol{E}, Z_p)$, pour tout p premier impair. Si $\bar{\boldsymbol{E}}$ et $\boldsymbol{Spin}(2n)$ sont les revêtements universels (à deux feuillets) de $\boldsymbol{E}$ et $\boldsymbol{SO}(2n)$, il en est alors de même de l'application correspondante de $H^*(\boldsymbol{Spin}(2n), Z_p)$ dans $H^*(\bar{\boldsymbol{E}}, Z_p)$; le Théorème 3 du Chap. III de [13] montre alors que $\pi_i(\bar{\boldsymbol{E}}) \to \pi_i(\mathbf{Spin}(2n))$, donc aussi $\pi_i(\boldsymbol{E}) \to \pi_i(\boldsymbol{SO}(2n))$, est un C-isomorphisme sur, ce qui termine la démonstration.

19. Groupes d'homotopie de dimension 6 des groupes classiques. Nous supposerons connues les valeurs des cinq premiers groupes d'homotopie des groupes classiques (*voir* [14], 24. 11, 25. 4, 25. 5, et [9], 3. 72) et le fait que $\pi_6(\boldsymbol{S}_3) \approx Z_{12}$.[7] Pour déterminer les 6-ièmes groupes d'homotopie, nous nous appuyerons sur la:

[7] Il est classique que $\pi_6(S_3)$ a 12 éléments; on en trouvera une démonstration simple dans [13], Chap. IV, Prop. 10. Pour prouver qu'il est cyclique, on peut, soit montrer qu'il contient un sous-groupe isomorphe à Z_4, ce qui a été fait par M. G. Barratt et G. F. Paechter, *Proc. Nat. Acad. Sci. U. S. A.*, tome 38 (1952), pp. 119-121, soit utiliser

PROPOSITION 19.1. *La classe caractéristique* $\alpha \in \pi_6(S_3)$ *de la fibration* $Sp(2)/Sp(1)$ *est un générateur de* $\pi_6(S_3)$.

Nous savons déjà par 17.2(c) que α n'est pas divisible par trois; pour obtenir 19.1, il nous suffit donc de montrer que α n'est pas divisible par deux.

Identifions $Sp(1)$ à la sphère unité du corps des quaternions, et S_6 à la variété des couples (q, q') de quaternions tels que $|q|^2 + |q'|^2 = 1$ et que la partie réelle de q' soit nulle; d'après [14], 24.11, la classe α est alors définie par l'application $g: S_6 \to S_3$ qui vérifie:

$$g(q, q') = 1 - 2q \cdot (1 + q')^{-2} \cdot \bar{q}.$$

D'après G. W. Whitehead [17], l'application g est homotope à g':

$$g'(q, q') = 1 - 2\,|q|^2 + 2 \cdot \frac{q \cdot q' \cdot \bar{q}}{|q|}.$$

Si l'on pose $q = Q \cdot \cos\theta$, $q' = Q' \cdot \sin\theta$, avec $0 \leq \theta \leq \pi/2$ et $|Q| = |Q'| = 1$, on voit que

$$g'(Q, Q') = -\cos 2\theta + \sin 2\theta \cdot Q \cdot Q' \cdot \bar{Q}.$$

Cela signifie que $g': S_6 \to S_3$ est obtenue en faisant la construction de Hopf sur l'application de $S_3 \times S_2$ dans S_2 donnée par $(Q, Q') \to Q \cdot Q' \cdot \bar{Q}$. Mais alors, d'après Blakers-Massey, (*Proc. Nat. Acad. Sci., U. S. A.*, vol. 35, (1949), 322-328), l'image de α par l'invariant de Hopf généralisé $H: \pi_6(S_3) \to Z_2$ est *non nulle*; α n'est donc pas divisible par deux.

PROPOSITION 19.2. *On a* $\pi_6(Sp(1)) = Z_{12}$, *et* $\pi_6(Sp(n)) = 0$ *pour* $n \geq 2$.

La première égalité résulte de $Sp(1) = S_3$. La fibration $Sp(2)/Sp(1) = S_7$ donne lieu à la suite exacte:

$$\pi_7(S_7) \xrightarrow{d} \pi_6(S_3) \to \pi_6(Sp(2)) \to 0.$$

L'image de d étant le sous-groupe engendré par α est *tout* $\pi_6(S_3)$ d'après 19.1, il s'ensuit que $\pi_6(Sp(2)) = 0$, d'où $\pi_6(Sp(n)) = 0$ pour $n \geq 2$.

PROPOSITION 19.3. *On a* $\pi_6(SO(3)) = Z_{12}$, $\pi_6(SO(4)) = Z_{12} + Z_{12}$, $\pi_6(SO(n)) = 0$ *pour* $n \geq 5$.

$SO(3)$ et $SO(4)$ ont pour revêtements universels S_3 et $S_3 \times S_3$ respectivement, d'où les deux premiers résultats. Il est classique que le revêtement universel de $SO(5)$ est isomorphe à $Sp(2)$, donc, vu 19.1, $\pi_6(SO(5)) = 0$.

La fibration $SO(6)/SO(5) = S_5$ donne lieu à la suite exacte:

---

la détermination des groupes d'Eilenberg-MacLane en cohomologie modulo 2 due à l'un de nous, *C. R. Acad. Sci. Paris*, tome 234 (1952), pp. 1243-1245, ainsi qu'un article à paraître aux *Comm. Math. Helv.*

$$0 \to \pi_6(\boldsymbol{SO}(6)) \to \pi_6(\boldsymbol{S}_5) \xrightarrow{d} \pi_5(\boldsymbol{SO}(5)) \to \pi_5(\boldsymbol{SO}(6)).$$

Comme $\pi_5(\boldsymbol{SO}(6)) = Z$ et $\pi_5(\boldsymbol{SO}(5)) = Z_2$, (*voir* [9]), on voit que d est un isomorphisme de $\pi_6(\boldsymbol{S}_5) = Z_2$ sur $\pi_5(\boldsymbol{SO}(5))$, d'où $\pi_6(\boldsymbol{SO}(6)) = 0$.

Considérons maintenant la suite exacte:

$$0 \to \pi_6(\boldsymbol{SO}(7)) \to \pi_6(\boldsymbol{S}_6) \xrightarrow{d} \pi_5(\boldsymbol{SO}(6)) \to \pi_5(\boldsymbol{SO}(7)).$$

D'après [9], on a $\pi_5(\boldsymbol{SO}(7)) = 0$ et $\pi_5(\boldsymbol{SO}(6)) = Z$; par suite, d est un isomorphisme sur et $\pi_6(\boldsymbol{SO}(7)) = 0$; comme $\pi_6(\boldsymbol{SO}(n)) = \pi_6(\boldsymbol{SO}(7))$ pour $n \geq 7$, la démonstration de 19.3 est achevée.

PROPOSITION 19.4. *On a*

$$\pi_6(\boldsymbol{SU}(2)) = Z_{12},\ \pi_6(\boldsymbol{SU}(3)) = Z_6,\ \pi_6(\boldsymbol{SU}(n)) = 0 \qquad (n \geq 4).$$

La première égalité résulte de $\boldsymbol{SU}(2) = \boldsymbol{S}_3$. Examinons le groupe $\boldsymbol{SU}(3)$; la fibration $\boldsymbol{SU}(3)/\boldsymbol{S}_3 = \boldsymbol{S}_5$ donne lieu à la suite exacte:

$$\pi_7(\boldsymbol{S}_5) \xrightarrow{d} \pi_6(\boldsymbol{S}_3) \to \pi_6(\boldsymbol{SU}(3)) \to \pi_6(\boldsymbol{S}_5) \xrightarrow{d} \pi_5(\boldsymbol{S}_3) \to \pi_5(\boldsymbol{SU}(3)).$$

Montrons tout d'abord que l'homomorphisme $d: \pi_6(\boldsymbol{S}_5) \to \pi_5(\boldsymbol{S}_3)$ applique le premier groupe *sur* le second. On sait [11] que $\pi_4(\boldsymbol{SU}(3)) = 0$, donc qu'une application quelconque $\boldsymbol{S}_4 \to \boldsymbol{S}_3 \to \boldsymbol{SU}(3)$ est inessentielle; il en est a fortiori de même pour sa composée avec une application $\boldsymbol{S}_5 \to \boldsymbol{S}_4$; comme $\boldsymbol{S}_5 \to \boldsymbol{S}_4 \to \boldsymbol{S}_3$ est essentielle lorsque $\boldsymbol{S}_5 \to \boldsymbol{S}_4$ et $\boldsymbol{S}_4 \to \boldsymbol{S}_3$ le sont, cela implique que l'image de $\pi_5(\boldsymbol{S}_3)$ dans $\pi_5(\boldsymbol{SU}(3))$ est nulle, donc que d applique $\pi_6(\boldsymbol{S}_5)$ sur $\pi_5(\boldsymbol{S}_3)$. Par conséquent, la suite exacte précédente donne:

$$\pi_7(\boldsymbol{S}_5) \xrightarrow{d} \pi_6(\boldsymbol{S}_3) \xrightarrow{i} \pi_6(\boldsymbol{SU}(3)) \to 0.$$

Pour établir l'égalité $\pi_6(\boldsymbol{SU}(3)) = Z_6$, il suffit de montrer que l'image de d n'est pas nulle, autrement dit que le noyau de i est non nul. Or Hilton a prouvé que l'application composée $\boldsymbol{S}_6 \to \boldsymbol{S}_5 \to \boldsymbol{S}_4 \to \boldsymbol{S}_3$, où chaque application est essentielle, définit un élément non nul de $\pi_6(\boldsymbol{S}_3)$, (*voir* aussi [13], Chap. IV, Prop. 10, Rem. 1), et le raisonnement fait plus haut, (qui s'applique *a fortiori* ici), montre que cet élément est dans le noyau de i.

Il est bien connu que le groupe $\boldsymbol{SU}(4)$ est isomorphe au revêtement universel du groupe $\boldsymbol{SO}(6)$; on a donc $\pi_6(\boldsymbol{SU}(4)) = 0$ d'après 19.3 et comme $\pi_6(\boldsymbol{SU}(n)) = \pi_6(\boldsymbol{SU}(4))$ pour $n \geq 4$, la Proposition 19.4 est complètement démontrée.

20. Les fibrations des variétés de Stiefel complexes. Soient $\boldsymbol{W}_{n,q} = \boldsymbol{U}(n)/\boldsymbol{U}(n-q)$ la variété de Stiefel complexe des q-repères orthonormaux de l'espace hermitien C^n, et $\psi_{q,r}$ la projection naturelle de $\boldsymbol{U}(n)/\boldsymbol{U}(n-q) = \boldsymbol{W}_{n,q}$ sur $\boldsymbol{U}(n)/\boldsymbol{U}(n-r) = \boldsymbol{W}_{n,r}$ $(q \geq r)$. On sait ([2], § 9) que $H^*(\boldsymbol{W}_{n,q}, Z)$

est une algèbre extérieure engendrée par des éléments de dimensions $2n-2q+1,\ 2n-2q+3,\cdots,2n-1$, appliquée biunivoquement dans $H^*(\boldsymbol{U}(n),Z)$ par $\psi^*_{n,q}$, et que

20.1 $$H^*(\boldsymbol{U}(n),Z)\approx H^*(\boldsymbol{U}(n-q),Z)\otimes H^*(\boldsymbol{W}_{n,q},Z).$$

Plus précisément, si l'on désigne par $h_1,\cdots,h_n$ les générateurs universellement transgressifs de $H^*(\boldsymbol{U}(n),Z)$ définis au No. 4, on a:

LEMME 20.2. *L'image de $H^*(\boldsymbol{W}_{n,q},Z)$ dans $H^*(\boldsymbol{U}(n),Z)$ par $\psi^*_{n,q}$ est la sous-algèbre de $H^*(\boldsymbol{U}(n),Z)$ engendrée par les éléments h_i, $n-q+1\leq i\leq n$.*

Soit v_i un générateur de $H^{2n-2i+1}(\boldsymbol{W}_{n,i},Z)$ $(1\leq i\leq n)$. C'est un élément de dimension positive minimum de $H^*(\boldsymbol{W}_{n,i},Z)$, par conséquent, d'après un raisonnement aisé, exposé dans [2], § 23, l'élément $\psi^*_{n,i}(v_i)$ est universellement transgressif, d'où $\psi^*_{n,i}(v_i)=m_ih_i$; mais les éléments $\psi^*_{n,i}(v_i)$ forment un système de générateurs de $H^*(\boldsymbol{U}(n),Z)$, ([2], § 9, Remarque 2), donc $m_i=\pm 1$ $(1\leq i\leq n)$.

La relation de transitivité évidente $\psi^*_{n,i}=\psi^*_{n,q}\circ\psi^*_{q,i}$, $(1\leq i\leq q)$, montre alors que les éléments h_i $(n-q+1\leq i\leq n)$ sont dans l'image de $\psi^*_{n,q}$; ils engendrent forcément toute cette image d'après 20.1.

Ce lemme permet de ramener le calcul des puissances réduites dans $H^*(\boldsymbol{W}_{n,q})$ au calcul analogue dans $H^*(\boldsymbol{U}(n))$, que nous avons déjà fait, (*voir* 11.4). Nous voulons en tirer des conditions nécessaires pour l'existence d'une section dans la fibration de $\boldsymbol{W}_{n,r+s}$ par $\boldsymbol{W}_{n-r,s}$, de base $\boldsymbol{W}_{n,r}$.

PROPOSITION 20.3. *Si la fibration $\boldsymbol{W}_{n,r+s}/\boldsymbol{W}_{n-r,s}=\boldsymbol{W}_{n,r}$ admet une section, on a, pour tout p premier, $\mathcal{P}_p^k(h_i)=0$, ou, ce qui revient au même, $b_p^{k,j}\equiv 0 \bmod p$, lorsque i vérifie les inégalités:*

(1) $n-r-s<i\leq n-r$,

(2) $n-r<j\leq n$, *en posant* $j=i+k(p-1)$.

Soit h'_a l'élément de $H^*(\boldsymbol{W}_{n,r+s})$ vérifiant

$$\psi^*_{n,r+s}(h'_a)=h_a \qquad (n-r-s<a\leq n)$$

et soit de même $h''_b\in H^*(\boldsymbol{W}_{n,r})$ tel que

$$\psi^*_{n,r}(h''_b)=h_b \qquad (n-r<b\leq n).$$

On a évidemment $\psi^*_{r+s,r}(h''_b)=h'_b$, et si $s:\boldsymbol{W}_{n,r}\rightarrow\boldsymbol{W}_{n,r+s}$ est une section, l'égalité $s^*\circ\psi^*_{r+s,r}=1$, donne alors:

20.4 $$\begin{aligned} s^*(h'_i)&=0 && (n-r-s<i\leq n-r)\\ s^*(h'_j)&=h''_j && (n-r<j\leq n); \end{aligned}$$

si maintenant nous supposons que i vérifie les inégalités de 20. 3, on déduit de 11. 4, 20. 2 et 20. 4:

$$0 = \mathcal{P}_p^k(s^*(h'_i)) = s^*(\mathcal{P}_p^k(h'_i)) = s^*(b_p^{k,j} \cdot h'_j) = b_p^{k,j} \cdot s^*(h'_j) = b_p^{k,j} \cdot h''_j$$

donc $b_p^{k,j} \equiv 0$ *mod* p.

Corollaire 20. 5. *Si la fibration* $W_{n,r+s}/W_{n-r,s} = W_{n,r}$ *admet une section, on a* $s = 1$ *ou* $r = 1$.

Supposant $r \geq 2$ et $s \geq 2$, nous appliquerons la Proposition 20. 3 avec $p = 3$; distinguons deux cas:

a) $n - r \equiv 2 \bmod 3$; on pose $i = n - r$, $k = 1$; on a donc $j = i + 2 = n - r + s$ et les inégalités (1) et (2) de 20. 3 sont vérifiées puisque $r \geq 2$; de plus, vu l'hypothèse faite sur $n - r$, on a $j \equiv 1 \bmod 3$. Mais d'après 12. 3 $b_3^{1,j} \equiv j \bmod 3$, donc $b_3^{1,j} \not\equiv 0 \bmod 3$ et il n'y a pas de section vu 20. 3.

(b) $n - r \not\equiv 2 \bmod 3$. On pose $i = n - r - 1$, $k = 1$; les inégalités (1) et (2) sont vérifiées puisque $s \geq 2$; mais $j = n - r + 1 \not\equiv 0 \bmod 3$, d'où comme précédemment, $b_3^{1,j} \not\equiv 0 \bmod 3$.

Nous traiterons maintenant plus en détails le cas $r = 1$; on pourrait étudier de même le cas $s = 1$, mais, les calculs étant plus compliqués, nous ne nous y attarderons pas.

Proposition 20. 6. *Si la fibration* $W_{n,s+1}/W_{n-1,s} = W_{n,1} = S_{2n-1}$ *admet une section, n est divisible par l'entier*

$$N_s = \Pi_p \, p^{1+h(p,s)}$$

où le produit est étendu à les nombres premiers, et où $h(p,s)$ *désigne pour tout* p *le plus grand entier* $h \geq -1$ *tel que* $(p-1) \cdot p^h \leq s$.

Vu la Proposition 20. 3, il nous suffira de prouver ceci:

Lemme 20. 7. *Soient p un nombre premier, s un entier* $< n$, *et supposons que* $b_p^{k,n} \equiv 0$ *mod p pour tout entier* $k > 0$ *tel que* $n - k(p-1) \geq n - s$, *(autrement dit tel que* $k(p-1) \leq s$). *Alors n est divisible par* $p^{1+h(p,s)}$.

Ce lemme sera lui-même conséquence du lemme suivant, dans lequel s n'intervient plus:

Lemme 20. 8. *Soient a un entier* ≥ 0, $k = p^a$, *supposons* $n > k(p-1)$ *et n divisible par k. Si* $b_p^{k,n} \equiv 0$ *mod p, alors n est divisible par* p^{a+1}.

Montrons, par récurrence sur s, que 20. 8 entraîne 20. 7. Ce dernier est trivial pour $s = 0$, supposons le vrai pour $s - 1$; puisque

$$h(p,s) \leq 1 + h(p, s-1),$$

cela implique que $p^{h(p,s)}$ divise n; mais $(p-1)\cdot p^{h(p,s)} \leq s < n$ par définition de $h(p,s)$, par conséquent l'entier $a = p^{h(p,s)}$ vérifie les hypothèses de 20. 8 et n est bien divisible par $p^{h(p,s)+1}$.

Le lemme 20. 8 est un cas particulier du résultat suivant, que nous allons maintenant démontrer:

LEMME 20. 9. *Soient F le polynôme symétrique $\Sigma x_1^{\alpha_1}\cdots x_i^{\alpha_i}$, de degré $n = \Sigma \alpha_i$, p un nombre premier, a un entier. On suppose que p^a divise n et qu'il y a au plus p^a indices j tels que $\alpha_j \neq 1$.*

Dans ces conditions, pour que le terme dominant de F soit $\not\equiv 0 \bmod p$, il faut et il suffit, ou bien que $\alpha_j = 1$ pour tout j, ou bien qu'il y ait exactement p^a indices j tels que $\alpha_j \neq 1$, les α_j correspondants étant tous égaux et n n'étant pas divisible par p^{a+1}.

(On obtient 20. 8 en appliquant 20. 9 au cas où $\alpha_j = p$ pour $1 \leq j \leq p^a$ et $\alpha_j = 1$ pour $j > p^a$.)

Nous démontrerons le lemme 20. 9 par récurrence descendante sur i, le cas $i = n$ étant trivial puisque tous les α_j sont alors égaux à 1.

Nous supposerons que l'on a $\alpha_1, \cdots, \alpha_k > 1$, et $\alpha_{k+1} = \cdots = \alpha_i = 1$. On a $1 \leq k \leq p^a$, vu $i < n$. Considérons le polynôme symétrique:

$$G = (\Sigma x_1^{\alpha_1-1}\cdots x_k^{\alpha_k-1})\cdot(\Sigma x_1\cdots x_i)\,;$$

dans le développement de G, nous trouverons évidemment le polynôme F; quant aux autres termes ce seront des polynômes symétriques de la forme:

$$c_{(\beta)}\Sigma x_1^{\beta_1}\cdots x_k^{\beta_k}\cdot x_{k+1}\cdots x_{i'},$$

où $i' = n + k - \Sigma \beta_j$, et où β_j est égal soit à α_j, soit à $\alpha_j - 1$, ce second cas se présentant pour au moins une valeur de i, (ce qui montre que $i' > i$); le coefficient $c_{(\beta)}$ est un entier.

G est le produit de deux polynômes symétriques, il est donc décomposable et l'égalité:

$$G = F + \Sigma_{(\beta)} c_{(\beta)}\Sigma x_1^{\beta_1}\cdots x_k^{\beta_k}\cdot x_{k+1}\cdots x_{i'},$$

montre que le terme dominant de F, changé de signe, est égal à la somme des termes dominants des polynômes $c_{(\beta)}\Sigma x_1^{\beta_1}\cdots x_k^{\beta_k}x_{k+1}\cdots x_{i'}$; comme $i' > i$, on peut appliquer le lemme à ces derniers vu l'hypothèse de récurrence. Nous distinguerons quatre cas:

(A) *On a $\alpha_j = 2$ pour $j \leq k$.* Dans ce cas, le seul choix des β_j qui conduise à un terme dominant non nul est celui où $\beta_j = 1$ pour tout j; il est immédiat que le coefficient $c_{(\beta)}$ correspondant est égal au coefficient binomial $\binom{n}{k}$, et, puisque $1 \leq k \leq p^a$ et que p^a divise n, les propriétés de divisibilité des coefficients binomiaux montrent que:

$$c_{(\beta)} = \binom{n}{k} \equiv 0 \bmod p \quad \text{si } k < p^a,$$

$$c_{(\beta)} = \binom{n}{k} \equiv n/p^a \bmod p \quad \text{si } k = p^a,$$

ce qui établit le lemme sous l'hypothèse (A).

(B) *On a* $\alpha_j = q > 2$ *pour* $j \leq k$. Le seul choix des β_j qui conduise à un terme dominant non nul est celui où $\beta_j = q - 1$ pour $1 \leq j \leq k$, lorsque de plus $k = p^a$ et p^{a+1} ne divise pas n; on voit alors tout de suite que $c_{(\beta)} = 1$, ce qui démontre le lemme dans le cas (B).

(C) *On a* $k < p^a$. Le seul choix des β_j qui conduise à un terme dominant non nul est celui où $\beta_j = 1$ pour tout j; mais les α_j, $(1 \leq j \leq k)$, sont alors égaux à 2, et l'on se retrouve dans le cas (A) déjà traité.

(D) *On a* $k = p^a$, *et les* α_j, $(1 \leq j \leq k)$, *ne sont pas tous égaux.* Dans ce cas, le seul choix des β_j qui conduise à un terme dominant non nul est $\beta_1 = \cdots = \beta_k = q > 1$, et ce choix n'est possible que si certains des α_j sont égaux à $q + 1$, et tous les autres à q; soient r et s leurs nombres respectifs. Vu l'hypothèse faite on a $r + s = p^a$, $r \geq 1$, $s \geq 1$; comme $q > 1$, il est immédiat que

$$c_{(\beta)} = \binom{r+s}{r} = \binom{p^a}{r} \equiv 0 \bmod p,$$

d'où le lemme dans le cas (D).

Comme les cas (A), (B), (C), (D) épuisent toutes les possibilités, la démonstration du lemme 20. 8 est achevée, et la Proposition 20. 6 est complètement établie.

Exemples.

$s = 1$; pour $p = 2$, on a $h(p, 1) = 0$, pour $p > 2$, $h(p, 1) = -1$, d'où $N_1 = 2$. Dans ce cas du reste, la condition "n pair" est non seulement nécessaire mais suffisante pour l'existence d'une section (*voir* Eckmann, [8], Satz IV).

$s = 2$; pour $p = 2$, $h(p, 2) = 1$, pour $p = 3$, $h(p, 2) = 0$, pour $p \geq 5$, $h(p, 2) = -1$, donc $N_s = 12$. Nous ignorons si la condition "n divisible par 12" est suffisante pour l'existence d'une section, mais cela semble peu probable. Il est assez naturel de conjecturer que $\boldsymbol{W}_{n,s+1}/\boldsymbol{W}_{n-1,s} = \boldsymbol{S}_{2n-1}$ n'a pas de section si $s > 1$.

Indiquons pour terminer quelques valeurs de la fonction arithmétique N_s:

$$N_1 = 2,\ N_2 = N_3 = 12,\ N_4 = N_5 = 120,\ N_6 = N_7 = 2.520, \cdots,$$

$$N_{20} = N_{21} = 6.983.776.800.$$

INSTITUTE FOR ADVANCED STUDY, PRINCETON,
UNIVERSITÉ DE NANCAGO.

BIBLIOGRAPHIE.

[1] A. Borel, "Le plan projectif des octaves et les sphères comme espaces homogènes," *C. R. Acad. Sci. Paris*, tome 230 (1950), pp. 1378-1380.

[2] ———, "Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts," *Annals of Mathematics*, tome 57 (1953), pp. 115-207.

[3] ———, "La cohomologie mod 2 de certains espaces homogènes," à paraître aux *Comm. Math. Helv.*

[4] ——— et J.-P. Serre, "Détermination des p-puissances réduites de Steenrod dans la cohomologie des groupes classiques. Applications," *C. R. Acad. Sci. Paris*, tome 233 (1951), pp. 680-682.

[5] H. Cartan et J.-P. Serre, "Espaces fibrés et groupes d'homotopie. I. Constructions générales," *C. R. Acad. Sci. Paris*, tome 234 (1952), pp. 288-290; "II. Applications," *ibid.*, pp. 393-395.

[6] S. S. Chern, "Characteristic classes of hermitian manifolds," *Annals of Mathematics*, tome 47 (1946), pp. 85-121.

[7] ———, "On the characteristic classes of complex spheres bundles and algebraic varieties," à paraître dans l'*Amer. Jour. Math.*

[8] B. Eckmann, "Systeme von Richtungsfeldern in Sphären und stetige Lösungen komplexer linearer Gleichungen," *Comm. Math. Helv.*, tome 15 (1942), pp. 1-26.

[9] ———, "Espaces fibrés et homotopie," *Colloque de Topologie*, Bruxelles, 1950, pp. 83-89.

[10] C. E. Miller, "The topology of rotation groups," *Annals of Mathematics*, tome 57 (1953), pp. 90-114.

[11] L. S. Pontrjagin, "Ueber die topologische Struktur der Lie'schen Gruppen," *Comm. Math. Helv.*, tome 13 (1940), pp. 277-283.

[12] J.-P. Serre, "Homologie singulière des espaces fibrés. Applications," *Annals of Mathematics*, tome 54 (1951), pp. 425-505.

[13] ———, "Groupes d'homotopie et classes de groupes abéliens," à paraître aux *Annals of Mathematics*.

[14] N. E. Steenrod, *The topology of fibre bundles*, Princeton, 1951.

[15] ———, a. "Homology groups of symmetric groups and reduced powers operations" *Proc. Nat. Acad. Sci. U. S. A.*, tome 39 (1953), pp. 213-217, b. "Cyclic reduced powers of cohomology classes" *ibid.*, pp. 217-223.

[16] R. Thom, "Une théorie axiomatique des puissances de Steenrod," *Colloque de Topologie*, Strasbourg, 1951.

[17] G. W. Whitehead, "Correction to my paper 'On families of continuous vector fields over spheres,'" *Annals of Mathematics*, tome 48 (1947), pp. 782-783.

[18] W. T. Wu, "Sur la structure presque complexe d'une variété différentiable réelle," *C. R. Acad. Sci. Paris*, tome 228 (1949), pp. 972-973.

[19] ———, *Sur les classes caractéristiques des structures fibrées sphériques*, *Act. Sci. et Ind.* 1183, Hermann éd., Paris (1952).

[20] ———, "Sur les puissances de Steenrod," *Colloque de Topologie*, Strassbourg, 1951.

ON THE LOCAL BEHAVIOR OF SOLUTIONS OF NON-PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

This paper deals with local problems concerning binary elliptic systems of partial differential equations in two independent variables and with corresponding partial elliptic differential equations of second order. The results are known in the analytic case, but most of them are not known even in the C^∞-case. The transition to non-analytic cases will be made possible by appropriate applications of the ideas contained in a short paper of Carleman [1]. At the same time, Carleman's results will be made more precise.

At the end of the paper, corresponding extensions and applications will be made of the results of Carleman [2] on non-parabolic systems in n dependent and two independent variables.

1. The gradient. For a function u of (x, y), let p, q and r, s, t denote, as usual, the first and second order partial derivatives. In what follows, all partial differential equations are assumed to involve only real-valued coefficients and/or functions, and all solutions are assumed to be real-valued, unless the contrary is said or implied.

THEOREM 1. *Let $d(x,y)$, $e(x,y)$, $f(x,y)$ be continuous functions on a circle*

$$x^2 + y^2 < R^2. \tag{1}$$

Let $u = u(x,y)$ be a function, of class C^2 on (1), satisfying the partial differential equation

$$r + t + dp + eq + fu = 0 \tag{2}$$

or, more generally, let $u(x,y)$ be a function, of class C^1 on (1), satisfying the integral relation

$$\int_J qdx - pdy = \iint_E (dp + eq + fu)\,dxdy \tag{3}$$

for every domain E bounded by a piecewise smooth (C^1) Jordan curve J

* Received January 9, 1953.

contained in (1). *Then, if the solution u behaves, as $(x,y) \to (0,0)$, in such a way that*

$$u(x,y) = o(\rho^n), \quad \text{where} \quad \rho = (x^2+y^2)^{\frac{1}{2}}, \tag{4}$$

holds for some non-negative integer n, the gradient (u_x, u_y) of u must behave in such a way that

$$\lim_{\rho \to 0} (u_y + iu_x)/(x+iy)^n \text{ exists} \tag{5}$$

(*for the same $n \geqq 0$*).

If the limit in (5) is denoted by $c_2 + ic_1$, where c_1, c_2 are real constants, then the assertion (5) can be written as

$$\begin{aligned} u_x &= \rho^n (c_1 \cos n\theta + c_2 \sin n\theta) + o(\rho^n), \\ u_y &= \rho^n (c_2 \cos n\theta - c_1 \sin n\theta) + o(\rho^n), \end{aligned} \tag{5'}$$

where $x = \rho \cos \theta$, $y = \rho \sin \theta$, which implies that the assumption (4) can be refined to

$$u = (n+1)^{-1} \rho^{n+1} \{c_1 \cos(n+1)\theta + c_2 \sin(n+1)\theta\} + o(\rho^{n+1}). \tag{5''}$$

In fact, (5″) follows if (5′) is inserted into the identity

$$u = \int_0^\rho (p \cos \theta + q \sin \theta)\, d\rho,$$

in which the argument of u, p, q is $(\rho \cos \theta, \rho \sin \theta)$.

The content of the following theorem is that, unless $u \equiv 0$, the relation (4) cannot hold for every n; so that (5′), (5″) hold for some $n \geqq 0$ with constants c_1, c_2 which are not both 0.

THEOREM 2. *In addition to the conditions of Theorem* 1, *assume that $u \not\equiv 0$. Then there exists a least integer $m (> 0)$ such that* (4) *fails to hold for $n = m$ (and, therefore, the limit in* (5) *is not* 0 *when $n = m - 1$).*

An immediate consequence of Theorems 1 and 2 is the following:

COROLLARY 1. *Under the assumptions of Theorems* 1 *and* 2, *the zeros of* grad $u = (u_x, u_y)$ *cannot cluster at* $(x,y) = (0,0)$.

REMARK. It will be clear from the proofs that Theorems 1 and 2 remain unaltered if $dp + eq + fu$ in (2) or (3) is replaced by a more general function of (x, y, u, p, q), say $g = g(x, y, u, p, q)$, which is continuous in its

five variables and is subject to the restriction that, for every $\epsilon > 0$ and every $M > 0$, there exists a constant $K = K(\epsilon, M)$ satisfying

$$|g(x, y, u, p, q)| \leqq K \cdot (|u| + |p| + |q|)$$

when $x^2 + y^2 < \epsilon^2$ and $|u| \leqq M$, $|p| \leqq M$, $|q| \leqq M$. Similar remarks apply to the theorems below dealing with linear differential equations.

It is seen from (3) and the integrability condition

$$\int_J pdx + qdy = 0 \tag{6}$$

that, if $f \equiv 0$ and if u is of class C^2, then the functions p, q are of class C^1 and satisfy a system of partial differential equations of the type considered by Carleman [1]. Hence, in this case, the first (non-parenthetical) part of Theorem 2 and Corollary 1 are contained in Carleman's results. On the other hand, Theorem 1 and the last (parenthetical) part of Theorem 2 go beyond the results of Carleman and supply the asymptotic formulae (5′), (5″).

In the note referred to, Carleman states, without proof, a theorem on non-linear elliptic equations which implies (whether or not $f \equiv 0$) a weaker conclusion than that of Corollary 1, namely, that the *common zeros of* grad u *and of* u cannot cluster at $(x, y) = (0, 0)$. It turns out that Theorems 1 and 2 imply that the conclusion of this theorem of Carleman can be strengthened along the lines of the conclusion of Corollary 1; cf. Section **12** below.

2. Counterexamples. It is instructive to contrast the cases $f \equiv 0$ and $f \not\equiv 0$. If $f \equiv 0$, then the assumption (4) can be replaced by

$$u(x, y) = u(0, 0) + o(\rho^n) \tag{7}$$

without invalidating the conclusion (5) of Theorem 1, the corresponding conclusion of Theorem 2 (namely, that (7) cannot hold for every n unless $u \equiv u(0, 0)$) and the assertion of Corollary 1. If, on the other hand, $f \not\equiv 0$ and $u(0, 0) \neq 0$, then none of these conclusions holds; in other words, it is not true that (7) implies (5), nor is it true that $u \equiv$ const. if (7) holds for every n; finally, it is not true that $u \equiv$ const. when the zeros of grad u cluster at $(x, y) = (0, 0)$. With regard to the last negative assertion, the situation is the same as in the case of an *ordinary*, homogeneous linear differential equation

$$u_{xx} + f(x)u = 0, \qquad u = u(x). \tag{8}$$

The positive statements (concerning the case $f \equiv 0$) will be clear from

the proofs of Theorems 1 and 2. The negative statements (concerning the case $f \not\equiv 0$) can be verified by simple examples of (2) in which $d \equiv e \equiv 0$, and f is a function of x alone. In this case, any solution $u = u(x)$ of (8), when considered as a function of (x, y), is a solution of (2). The desired examples can be obtained by starting with a suitable function $u = u(x)$, where $u(x) \neq 0$, and defining $f(x)$ by (8). Thus, the example $u(x) = 1 + x^5 \sin(x^{-1})$, where $x \neq 0$ and $u(0) = 1$, proves the first and the third of the negative assertions (which state that the corresponding conclusions of Theorem 1 and Corollary 1 do not hold). The example $u(x) = 1 + \exp(-x^{-2})$, where $x \neq 0$ and $u(0) = 1$, proves the second negative assertion (which states that (7) can hold for every n even if $u \not\equiv \text{const.}$).

3. A binary elliptic linear system. The proofs of Theorems 1 and 2 will imply a proof of the following theorem (*), dealing with the integral analogue of a system of differential equations treated by Carleman [1]. His result is to the effect that (10) below cannot hold for every n (when u, v are of class C^1). The conclusion of (*) goes a great deal further. The consideration of the integral relations (9_1)-(9_2) below, instead of their differentiated form

$$\text{(9 bis)} \qquad v_x + u_y + \alpha u + \beta v = 0, \qquad v_y - u_x + \gamma u + \delta v = 0,$$

makes it possible to assume that u, v are only continuous.

THEOREM (*). *Let $\alpha(x,y)$, $\beta(x,y)$, $\gamma(x,y)$, $\delta(x,y)$ be continuous functions on (1). Let $u = u(x,y)$, $v = v(x,y)$ be continuous functions on (1) satisfying*

$$(9_1) \qquad \int_J u\,dx - v\,dy = \iint_E (\alpha u + \beta v)\,dx\,dy,$$

$$(9_2) \qquad \int_J v\,dx + u\,dy = \iint_E (\gamma u + \delta v)\,dx\,dy$$

for every domain E bounded by a piecewise smooth (C^1) Jordan curve contained in (1) (for instance, let u, v be of class C^1 and satisfy (9 bis)), and suppose that, for some non-negative integer n,

$$(10_n) \qquad u = o(\rho^n) \quad \text{and} \quad v = o(\rho^n)$$

as $\rho \to 0$, where $\rho = (x^2 + y^2)^{\frac{1}{2}}$. Then

$$(11_n) \qquad \lim_{\rho \to 0} (u + iv)/(x + iy)^{n+1} \text{ exists}$$

(for the same n). Moreover, unless both u and v vanish identically, there exists a least integer $n \geqq 0$ for which the limit (11_n) is not 0.

Theorem (*) will be clear from the proofs of Theorems 1 and 2. In fact, (9_1)-(9_2) imply that

$$\int_J (u+iv)(dx+idy) = \iint_E \{(\alpha+i\gamma)u+(\beta+i\delta)v\}dxdy,$$

and this relation can be used in the same way as (14) is used below in the proofs of Theorems 1 and 2.

4. Removable singularities. The Green formulae to be applied in the proof of Theorem 1 can be used to prove theorems generalizing the classical facts on the removable singularities of harmonic functions. In this direction there will be proved the following theorem (which can be considered to be the case $n=-1$ of Theorem (*)):

THEOREM ($\square$). *Let the coefficient functions of a system* (9 bis) *be continuous on the circle* (1) *and let u, v be functions of class C^1 on the punctured circle $0<x^2+y^2<R^2$, on which they satisfy* (9 bis); *while, as $x^2+y^2\to 0$, let*

(10_{-1}) $$u=o(1/\rho) \quad \text{and} \quad v=o(1/\rho),$$

where $\rho^2=x^2+y^2$. Then

(I) *u, v can be defined at $(x,y)=(0,0)$ so as to render u, v continuous at $(x,y)=(0,0)$; that is, the limits*

(11_{-1}) $$\lim_{\rho\to 0} u(x,y) \quad \text{and} \quad \lim_{\rho\to 0} v(x,y) \text{ exist;}$$

(II) *if, in addition, both limits (11_{-1}) are 0 and, correspondingly,*

(10_0) $$u(0,0)=0 \quad \text{and} \quad v(0,0)=0,$$

then the partial derivatives u_x, u_y, v_x, v_y exist, and (9 bis) *is satisfied, at $(x,y)=(0,0)$.*

Note that (II) does not claim the continuity of the partial derivatives at $(0,0)$; that is, the C^1-character of the solution on (1).

If the additional condition (10_0) is omitted, then the assertion of (II) is false in general. In order to see this, let $g(x,y)$ be $x^2/\{(x^2+y^2)\log(x^2+y^2)\}$ or 0 according as $\rho\neq 0$ or $\rho=0$. Then, according to Petrini [7], p. 138, the Poisson equation $z_{xx}+z_{yy}=g$ has no solution of class C^2 on the circle $x^2+y^2<R^2$ (for any $R<1$). But there exist functions $z=z(x,y)$ which are of class C^1 on the circle (1), are of class C^2 and satisfy the Poisson equation on the punctured circle $0<x^2+y^2<R^2$, while $z_{xx}(0,0)$ and

$z_{yy}(0,0)$ fail to exist. Since a harmonic function can be added to z, it can be supposed that $z_x(0,0) \neq 0$, hence, if $R > 0$ is small enough, $z_x \neq 0$ on (1). Then $u = z_y(x,y)$ and $v = z_x(x,y)$ are continuous on (1) and are of class C^1 and satisfy (9 bis) for $0 < x^2 + y^2 < R^2$, where $\alpha \equiv 0$, $\beta = -g(x,y)/z_x(x,y)$ and $\gamma \equiv \delta \equiv 0$. Nevertheless, $u_y(0,0)$ and $v_x(0,0)$ fail to exist.

On the other hand, the additional condition (10_0) can be omitted in (II) if additional conditions are imposed on the coefficient functions α, β, γ, δ:

THEOREM (□ bis). *Let the coefficient functions of* (9 bis) *satisfy a uniform Hölder condition* (*of some order* λ, *where* $0 < \lambda < 1$) *in the circle* (1), *and let* u, v *satisfy the conditions of part* (I) *of Theorem* (□). *Then* u, v *are of class* C^1 *on* (1) (and their partial derivatives u_x, u_y, v_x, v_y satisfy, on every compact subset of (1), a uniform Hölder condition of every order $\mu < \lambda$).

It will be clear from the proof that, in Theorems (□) and (□ bis), the condition that u, v are of class C^1 and satisfy (9 bis) on $0 < x^2 + y^2 < R^2$ can be reduced to the assumption that u, v are continuous and satisfy (9_1), (9_2) for every simply connected subdomain E of $0 < x^2 + y^2 < R^2$, bounded by a piecewise smooth (C^1) Jordan curve contained in $0 < x^2 + y^2 < R^2$.

It will also be clear from the proofs that Theorems (□) and (□ bis) have analogues in which the first order system (9 bis) is replaced by the second order equation (2).

REMARK TO THEOREM (□). The proof of Theorem (□) can be modified so as to show that if d, e, f are continuous on (1) and if, on the punctured circle $0 < x^2 + y^2 < R^2$, the function $u = u(x,y)$ is of class C^2, satisfies (2) and has partial derivatives p, q satisfying $p^2 + q^2 = o(1/\rho)$, then u can be defined at $(x,y) = (0,0)$ so as to become of class C^1 on (1). Also, if $u(0,0) = p(0,0) = q(0,0)$, then u_{xx}, $u_{xy} = u_{yx}$, u_{yy} exist, and (2) is satisfied, at $(x,y) = (0,0)$.

REMARK TO THEOREM (□ bis). The proof of Theorem (□ bis) will depend on a modification of a device of Lichtenstein [12], p. 1321, which implies that if u is continuous on (1) and satisfies the integral identity (3) in $E = E(J)$, then u is of class C^1 (and its partial derivatives satisfy, on every compact subset of (1), a uniform Hölder condition of every order $\mu < 1$) when d, e, f are continuous on (1); and that u is of class C^2 (and that its second order partial derivatives satisfy, on every compact subset of (1), a

uniform Hölder condition of every order $\mu < \lambda$) when d, e, f satisfy a uniform Hölder condition of order λ on (1); cf. [14], p. 735.

5. Proof of Theorem 1. Put $z = x + iy$, use the notation $g(z)$ for any function $g(x, y)$ of (x, y) (so that $g_x(z)$, $g_y(z)$ mean $\partial g(x, y)/\partial x$, $\partial g(x, y)/\partial y$, respectively), and let

(12) $\quad w = u_y + iu_x,$ $\qquad$ (13) $\quad W = du_x + eu_y + fu.$

Then (3) and the integrability condition (6) can be written as

$$\int_J w dz = \int\int_E W dxdy. \tag{14}$$

According to the Lemma of [5], p. 761, the identity (14) in $E = E(J)$ implies that

$$\int_J gwdz = \int\int_E \{gW + iw(g_x + ig_y)\} dxdy,$$

if $g = g(x, y)$ is any function of class C^1 on $E + J$. Let this be applied to the function $g(z) = z^{-k}(z - \zeta)^{-1}$, where $\zeta \neq 0$, k is a non-negative integer and $E = E_\epsilon$ is the domain bounded by the circles $|z| = R$, $|z| = \epsilon$ and $|z - \zeta| = \epsilon$ (where $\epsilon > 0$ is small and R is any fixed positive number smaller than the R in (1)). Since $g(z)$ is a regular function on E, hence $g_x + ig_y \equiv 0$, there results the identity

$$\int_J wz^{-k}(z - \zeta)^{-1} dz = \int\int_E Wz^{-k}(z - \zeta)^{-1} dxdy. \tag{15}$$

If $w(z)$ satisfies

$$w = o(|z|^{k-1}) \tag{16_k}$$

as $z \to 0$, then it follows from (15), on letting $\epsilon \to 0$, that

$$2\pi i w(\zeta)\zeta^{-k} = \int_{|z|=R} wz^{-k}(z - \zeta)^{-1} dz - \int\int_{|z|<R} Wz^{-k}(z - \zeta)^{-1} dxdy, \tag{17_k}$$

where the double integral is absolutely convergent.

Since Theorem 1 is trivial if $0 \leqq n \leqq 1$, it will be supposed that $n > 1$. It will be shown by an induction (on k for fixed n) that (4) implies (16_n). Clearly, $n > 1$ in (4) implies (16_1). Suppose that (16_k) holds for a k satisfying $1 \leqq k < n$. It will be verified that (16_{k+1}) holds.

The definition (13) shows that

$$|W| \leqq K(|w| + |u|) \tag{18}$$

holds for some constant $K = K_R > 0$. Hence

$$(19_k) \qquad 2\pi \,|\, w(\zeta)\zeta^{-k} \,| \leqq \int_{|z|=R} |\, wz^{-k}(z-\zeta)^{-1} \,| \cdot |\, dz \,| \\ + K \iint_{|z|<R} (|\, w \,| + |\, u \,|) \,|\, z \,|^{-k} \,|\, z - \zeta \,|^{-1} dxdy.$$

Let this inequality be multiplied by $|\, \zeta - z_0 \,|^{-1} d\xi d\eta$ and then integrated over $|\, \zeta \,| < R$, where $\zeta = \xi + i\eta$. Then, if use is made of the inequality (cf. [1], p. 473)

$$(20) \qquad \iint_{|z|<R} |\, z - \zeta \,|^{-1} dxdy < 2R, \quad \text{where} \quad |\, \zeta \,| < R,$$

and of the identity

$$(21) \qquad |\, (z-\zeta)(\zeta - z_0) \,|^{-1} = |\, z - z_0 \,|^{-1} \,|\, (z-\zeta)^{-1} + (\zeta - z_0)^{-1} \,|,$$

it follows that

$$2\pi \iint_{|z|<R} |\, w(z) z^{-k}(z-\zeta)^{-1} \,|\, dxdy \leqq 4R \int_{|z|=R} |\, wz^{-k}(z-\zeta)^{-1} \,| \cdot |\, dz \,| \\ + 4KR \iint_{|z|<R} (|\, w \,| + |\, u \,|) \,|\, z \,|^{-k} \,|\, z - \zeta \,|^{-1} \, dxdy,$$

where ζ has been written in place of z_0. Accordingly,

$$(2\pi - 4KR) \iint_{|z|<R} |\, wz^{-k}(z-\zeta)^{-1} \,|\, dxdy \\ \leqq 4R \int_{|z|=R} |\, wz^{-k}(z-\zeta)^{-1} \,| \cdot |\, dz \,| + 4KR \iint_{|z|<R} |\, uz^{-k}(z-\zeta)^{-1} \,|\, dxdy.$$

It can be supposed that R is so small that $2\pi - 4KR > 0$. Since $k < n$, hence

$$(22) \qquad uz^{-k} = O(1)$$

as $z \to 0$, the last double integral is $O(1)$ as $\zeta \to 0$. It follows therefore from (19_k) that $w(\zeta)\zeta^{-k} = O(1)$.

Thus, $W(z)z^{-k} = O(1)$ as $z \to 0$. This implies that the double integral in (17_k) tends to a limit as $\zeta \to 0$ (in fact, the absolute continuity of the set function $\iint_E Wz^{-k}(z-\zeta)^{-1}dxdy$ of E is uniform with respect to ζ, as $\zeta \to 0$; cf. (20)). Hence, by (17_k),

$$(23_k) \qquad \lim_{z \to 0} w(z) z^{-k} \text{ exists.}$$

If $k < n$, then the limit (23_k) is 0; for otherwise $u \neq o(|z|^{k+1})$, which contradicts (4), since $k+1 \leqq n$. Since this completes the induction, (16_n) holds for every n.

On the other hand, (23_k) was deduced as a consequence of (16_k) and (22). Hence (23_n), that is, (5), is true. This proves Theorem 1.

6. Proof of Theorem 2. Suppose that (4) holds for every positive integer n. It will be shown that $u(z) \equiv 0$ on (1). According to the proof of Theorem 1, (4) implies (16_n), which in turn implies (19_n). In view of (20), an integration of (19_n) leads to the inequality

$$2\pi \iint\limits_{|z|<R} |wz^{-n}|\, dxdy \leqq 2R \int\limits_{|z|=R} |wz^{-n}| \cdot |dz| + 2KR \iint\limits_{|z|<R} (|w| + |u|)|z|^{-n} dxdy,$$

which can be written as

$$(2\pi - 2KR) \iint\limits_{|z|<R} |wz^{-n}|\, dxdy \leqq 2R \int\limits_{|z|=R} |wz^{-n}| \cdot |dz| + 2KR \iint\limits_{|z|<R} |u|\, |z|^{-n} dxdy \tag{24}$$

(this inequality corresponds to (4) in [1], p. 473).

In order to appraise the last integral, note that

$$u(z) = \int_0^1 (u_x(tz)x + u_y(tz)y)\,dt; \text{ so that } |u(z)| \leqq \int_0^1 |zw(tz)|\,dt,$$

by (12). If this inequality is multiplied by $|z|^{-n}$, an integration gives

$$\iint\limits_{|z|<R} |uz^{-n}|\, dxdy \leqq \int_0^1 \{ \iint\limits_{|z|<R} |z^{-n+1}w(tz)|\, dxdy\} dt.$$

The last (triple) integral is transformed into

$$\int_0^1 t^{n-3} \{ \iint\limits_{|z|<tR} |z^{-n-1}w(z)|\, dxdy\} dt$$

by the change of variables $tz \to z$ in the interior double integral. If $n \geqq 3$, there results the inequality

$$\int_{|z|<R} | uz^{-n} | \, dxdy \leqq \int\int_{|z|<R} | wz^{-n+1} | \, dxdy.$$

Here the factor z^{-n+1} can be replaced by z^{-n} if it is supposed that $R < 1$ Thus it follows from (24) that

$$(25) \qquad (2\pi - 4KR) \int\int_{|z|<R} | wz^{-n} | \, dxdy \leqq 2R \int_{|z|=R} | wz^{-n} | \cdot | dz |$$

for $n = 3, 4, \cdots$.

Let $R(>0)$ be fixed and so small that $R < 1$ and $2\pi - 4KR > 0$. Suppose, if possible, that there exists a point $z = z_0$ in $|z| < R$ for which $w(z_0) \neq 0$. It is clear that the left side of (25) exceeds const. $|z_0|^{-n}$ for some const. > 0 which is independent of n. On the other hand, the right side of (25) is majorized by Const. R^{-n}. In view of $|z_0| < R$, this contradicts (25) for large n. Hence $w(z) \equiv 0$, and so $u(z) \equiv u(0) = 0$, for $|z| < R$.

Accordingly, the assumption of (4) for $n = 1, 2, \cdots$ implies that $u(z) \equiv 0$ holds on a circle $|z| < R$, where R is subject only to the restriction $R < \max(1, \frac{1}{2}\pi/K)$. Consequently, $u(z) \equiv 0$ holds on (1) without any restriction on R. This proves Theorem 2.

7. Proof of (I) **in Theorem** ($\square$). In this proof, the letter R will denote a positive number smaller than the R in (1). Let

$$(26) \qquad w = u + iv.$$

Then $w = w(x, y) = w(z)$ is $o(|z|^{-1})$ as $z \to 0$, by (10_{-1}). Hence, if $\zeta \neq 0$,

$$(27) \qquad 2\pi i w(\zeta) = \int_{|z|=R} w(z)/(z - \zeta) dz - \int\int_{|z|<R} (\alpha^* u + \beta^* v)/(z - \zeta) \, dxdy,$$

where $\alpha^* = \alpha + i\gamma$ and $\beta^* = \beta + i\delta$; cf. the proof (17_k). In fact, condition (10_{-1}) implies that, in the derivation of (27), the contributions of the line integral over $|z| = \epsilon$ and of the double integral over $|z| < \epsilon$ tend to 0 as $\epsilon \to 0$ (the double integral in (27) is absolutely convergent). If $K = K(R)$ is a constant such that $|\alpha^*|$ and $|\beta^*|$ are majorized by K for $|z| < R$, then (27) shows that

$$2\pi \, | w(\zeta) | \leqq \int_{|z|=R} | w(z)/(z - \zeta) | \cdot | dz | + K \int\int_{|z|<R} | w(z)/(z - \zeta) | \, dxdy.$$

If this inequality is multiplied by $d\xi d\eta / | \zeta - z_0 |$, an integration and (20), (21) lead to

$$(2\pi - 4KR) \iint_{|z|<R} | w(z)/(z-\zeta)| \, dxdy \leqq 4R \int_{|z|=R} | w(z)/(z-\zeta)| \cdot | dz |,$$

if ζ is written in place of z_0. If R is so small that $2\pi - 2KR > 0$, then the last two formula lines show that $w(\zeta)$ remains bounded as $\zeta \to 0$; cf. the proof of (16_k). Consequently, (27) implies that $w(\zeta)$ tends to a limit as $\zeta \to 0$; cf. the argument leading to (23_k). This proves (I).

PROOF OF (II) IN THEOREM ($\square$). The assumption (10_0) means that (26) satisfies $w = o(1)$ as $\zeta \to 0$. Hence w satisfies the identity

$$2\pi i w(\zeta)/\zeta = \int_{|z|=R} w(z)/(z-\zeta) dz - \iint_{|z|<R} (\alpha^* u + \beta^* v)/z(z-\zeta) dxdy;$$

cf. the proof of (17_k). The above arguments show, first, that $w(\zeta)/\zeta$ is bounded, and then, that $\lim w(\zeta)/\zeta$ exists, as $\zeta \to 0$. Hence the functions u, v possess partial derivatives at $(x, y) = (0, 0)$, and these satisfy $v_x + u_y = 0$, $v_y - u_x = 0$. But this means that (9 bis) holds at the point $(x, y) = (0, 0)$, since u and v vanish there.

PROOF OF THEOREM ($\square$ bis). It follows from the preceding two proofs that (26) satisfies (27). Since

$$(28) \qquad \iint_{|z|<R} dxdy / | z(z-\zeta)| = O(| \log | \zeta | |) \text{ as } \zeta \to 0,$$

it follows that if $0 < R' < R$, there exists a constant $K = K(R')$ satisfying

$$| w(\zeta_1) - w(\zeta_2) | \leqq K | \zeta_2 - \zeta_1 | | \log | \zeta_2 - \zeta_1 | |$$

if $| \zeta_1 | \leqq R'$, $| \zeta_2 | \leqq R'$.

In particular, u and v satisfy on $| z | \leqq R'$ a uniform Hölder condition of every order $\mu < 1$. Since α, β, γ, δ satisfy a uniform Hölder condition of order λ, it follows that $\alpha^* u + \beta^* v$ satisfies a uniform Hölder condition of order μ. The formula (27) is similar to that for the first derivatives of logarithmic potentials, with $\alpha^* u + \beta^* v$ having a rôle analogous to that of the density. Hence, w has partial derivatives satisfying a uniform Hölder condition of every order $\mu < \lambda$ on $| z | \leqq R'$. This proves Theorem ($\square$ bis).

8. The general linear case. The results of Section 2 can be trans-

cribed from (2) to the case of the general elliptic, homogeneous, linear differential equation,

$$ar + 2bs + ct + dp + eq + fu = 0, \tag{29}$$

where

$$ac - b^2 > 0 \qquad (\text{and } a > 0), \tag{30}$$

provided that the functions d, e, f of (x, y) are, as above, just continuous, but the functions a, b, c are of class C^1 on (1) (as to the parenthetical normalization in (30), note that (29) can be multiplied by -1). Because of the C^1-assumption, (29) can be written in the form

$$(ap + bq)_x + (bp + cq)_y + dp + eq + fu = 0, \tag{31}$$

where d, e, f are continuous functions (not identical with the corresponding functions in (29)). In order to consider "solutions u of class C^1 of (31)," it is convenient to write (31) as an integral identity,

$$\int_J (bp + cq)\,dx - (ap + bq)\,dy = \int_E\int (dp + eq + fu)\,dxdy, \tag{32}$$

where E is an arbitrary domain bounded by a piecewise smooth (C^1) Jordan curve J contained in (1).

Since a, b, c are of class C^1, it follows from the Lemma of [5], p. 761, that (32) is equivalent to

$$\int_J \Delta^{-1}\{(bp + cq)\,dx - (ap + bq)\,dy\} = \int_E\int (dp + eq + fu)\,dxdy, \tag{33}$$

where

$$\Delta = (ac - b^2)^{\frac{1}{2}} > 0, \tag{34}$$

and d, e, f represent another set of continuous functions. Consider the system of partial differential equations

$$\theta_x = (b\phi_x + c\phi_y)/\Delta, \qquad \theta_y = -(a\phi_x + b\phi_y)/\Delta \tag{35}$$

for the unknown functions θ, ϕ. According to Lichtenstein [11], the fact that a, b, c are of class C^1 (hence satisfy a uniform Hölder condition on every compact subset of (1)) implies that the system (35) has a solution

$$\theta = \theta(x, y), \qquad \phi = \phi(x, y) \tag{36}$$

of class C^1 on (1) such that (36) is a topological mapping of (1) onto a (θ, ϕ)-domain D^* and has a non-vanishing Jacobian $j = \partial(\theta, \phi)/\partial(x, y)$. Actually, $j > 0$, since

$$j = a\theta_x^2 + 2b\theta_x\theta_y + c\theta_y^2 = a\phi_x^2 + 2b\phi_x\phi_y + c\phi_y^2 > 0. \tag{37}$$

In terms of a C^1-solution $u = u(x, y)$ of (33) and the transformation $x = x(\theta, \phi)$, $y = y(\theta, \phi)$ inverse to (36), define a function $U(\theta, \phi)$ on D^* by

$$U(\theta, \phi) \equiv u(x, y). \tag{38}$$

Then U is of class C^1 and, according to (33) and (34), satisfies

$$\int_J U_\phi d\theta - U_\theta d\phi = \int\int_E (d^* U_\theta + e^* U_\phi + f^* U)\, d\theta d\phi, \tag{39}$$

where $d^* = (d\theta_x + e\theta_y)j$, $e^* = (d\phi_x + e\phi_y)j$, $f^* = fj$ are continuous functions of (θ, ϕ) and E is an arbitrary domain bounded by a piecewise smooth (C^1) Jordan curve J contained in D^*. The equation (39) is of the type (3).

REMARK. Note that even if $u(x, y)$ is a solution of class C^2 of (29), the function $U(\theta, \phi)$ need not be of class C^2. In fact, the transformation (36) need not be of class C^2 when it is only assumed that a, b, c are of class C^1; cf. [7], p. 265. Hence, even when starting with differential equations, (2) or (29), one is forced to consider the corresponding integro-differential equations, (3) or (32).

Any pair of functions $\theta^*(x, y)$, $\phi^*(x, y)$ for which $\theta^* + i\phi^*$ is a regular analytic function of $\theta + i\phi$ is a solution of (35). Hence, it can be supposed that θ and ϕ satisfy assigned "initial conditions" (consistent with (35)) at $(x, y) = (0, 0)$; for example,

$$\theta = \phi = 0 \text{ at } (x, y) = (0, 0) \tag{40}$$

and

$$\theta_x = \Delta,\ \theta_y = 0 \text{ and } \phi_x = -b,\ \theta_y = a \text{ at } (x, y) = (0, 0). \tag{41}$$

In this case,

$$\theta + i\phi \sim \Delta^0 x + i(a^0 y - b^0 x) \tag{42}$$

as $(x, y) \to (0, 0)$, where $a^0 = a(0, 0)$, $b^0 = b(0, 0), \cdots$ and

$$(U_\phi + iU_\theta)\Delta/a \sim \Delta^0 u_y + i(a^0 u_y - b^0 u_x). \tag{43}$$

Thus Theorems 1 and 2 can be transcribed as follows:

THEOREM 1*. *Let $a(x, y)$, $b(x, y)$, $c(x, y)$ be functions of class C^1 on* (1), *satisfying* (30), *hence, without loss of generality, the normalizations*

$$a(0, 0) = c(0, 0) = 1, \qquad b(0, 0) = 0, \tag{44}$$

and let $d(x, y)$, $e(x, y)$, $f(x, y)$ be continuous functions on (1). *Let $u = u(x, y)$ be a solution of class C^2 of* (29) *or, more generally, a function*

of class C^1 satisfying (32) *for every domain E bounded by a piecewise smooth* (C^1) *Jordan curve J contained in* (1). *Then, if u satisfies* (4) *for some integer $n \geqq 0$, it must satisfy* (5) (*and therefore* (5′), (5″) *as well*).

The normalization (44) can always be accomplished by a linear transformation of the (x, y)-plane.

REMARK. We were unable to decide whether or not the assertions of Theorem 1*, and of Theorem 2* below, remain true if the coefficients of (29) are just continuous.

THEOREM 2*. *In addition to the conditions of Theorem* 1*, *assume that $u \not\equiv 0$. Then the conclusions of Theorem* 2 *are valid.*

Since $\operatorname{grad} U = (U_\theta, U_\phi)$ vanishes if and only if $\operatorname{grad} u = (u_x, u_y)$ vanishes at the corresponding point, the analogue of Corollary 1 holds.

COROLLARY 1*. *Under the assumptions of Theorems* 1* *and* 2*, *the conclusion of Corollary* 1 *is valid.*

9. The general elliptic case of a linear system. The application of a conformal mapping makes possible not only the generalization of Theorems 1 and 2, dealing with the equation (2), to Theorems 1* and 2*, dealing with the equation (29), but also a generalization of Theorem (*) to a theorem dealing with a general linear elliptic system:

THEOREM (**). *Let a, b, c, d be functions of class C^1 on* (1) *with the property that the matrix*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{45}$$

has non-real roots and, without loss of generality, is normalized by the conditions

$$a(0,0) = d(0,0) = 0, \qquad b(0,0) = -c(0,0) = 1, \tag{45 bis}$$

and let α, β, γ, δ be continuous functions on (1). *Let u, v be continuous functions on* (1) *satisfying*

$$\int_J (au + by)\,dx + u\,dy = \int\int_E (\alpha u + \beta v)\,dx\,dy, \tag{46_1}$$

$$\int_J (cu + dv)\,dx + v\,dy = \int\int_E (\gamma u + \delta v)\,dx\,dy \tag{46_2}$$

for every domain E bounded by a piecewise smooth (C^1) *Jordan curve J contained in* (1) (*which means that*

$$(46\text{ bis})\qquad u_x - (au + bv)_y = \alpha u + \beta v, \qquad v_x - (cu + dv)_y = \gamma u + \delta v,$$

if the six functions $a, \cdots, v$ are supposed to be of class C^1, for instance). *Then, if* (10_n) *holds for some integer $n \geqq 0$, the conclusions of Theorem* (*) *are valid.*

PROOF OF THEOREM (**). Introduce the matrix and vector notation

$$(47)\qquad B = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \text{ and } \sigma = \begin{pmatrix} u \\ v \end{pmatrix}.$$

Then (46_1)-(46_2) can be written as a vector equation,

$$(48)\qquad \int_J A\sigma dx + \sigma dy = \iint_E B\sigma dxdy.$$

The assumptions on the matrix (45) show that if $\lambda \pm i\mu \equiv \lambda(x, y) \pm i\mu(x, y)$ are the characteristic numbers of (45), then there exists a non-singular matrix $T = T(x, y)$, of class C^1 on (1), satisfying

$$(49)\qquad TAT^{-1} = \begin{pmatrix} \lambda & -\mu \\ \mu & \lambda \end{pmatrix}, \qquad \mu \neq 0,$$

and $\lambda(0, 0) = 0$, $\mu(0, 0) = 1$; in particular, λ and μ are of class C^1 on (1). In place of the dependent variable (vector) σ, introduce the new dependent variable

$$(50)\qquad \tau = \begin{pmatrix} U \\ V \end{pmatrix} = T\sigma.$$

Then (48) becomes

$$(51)\qquad \int_J AT^{-1}\tau dx + T^{-1}\tau dy = \iint_E BT^{-1}\tau dxdy.$$

Since T is of class C^1, the Lemma of [5], p. 761, shows that

$$(52)\qquad \int_J TAT^{-1}\tau dx + \tau dy = \iint_E \{TBT^{-1} + T_xAT^{-1} - T_yAT^{-1}\}\tau dxdy.$$

The first and second components of the vector on the left of (52) are the line integrals of $(\lambda U - \mu V)dx + Udy$ and $(\mu U + \lambda V) + Vdy$, respectively. Hence, if the second component is multiplied by i and added to the first, there results the scalar equation

$$(53) \qquad \int_J (U+iV)\{(\lambda+i\mu)dx+dy\} = \iint_E (\alpha^*U+\beta^*V)\,dxdy,$$

where α^*, β^* are (complex-valued) continuous functions of (x,y) on (1).

The system

$$(54) \qquad \theta_x = \lambda\theta_y - \mu\phi_y, \qquad \phi_x = \mu\theta_y + \lambda\phi_y$$

of partial differential equations in the unknown functions θ, ϕ can be written in the form

$$\theta_x = \mu^{-1}(\lambda\phi_x - (\lambda^2+\mu^2)\phi_y), \qquad \theta_y = \mu^{-1}(\phi_x - \lambda\phi_y).$$

The latter is of the Cauchy-Riemann-Beltrami type (35). Hence the theorem of Lichtenstein [11] implies the existence of a C^1-solution $\theta = \theta(x,y)$, $\phi = \phi(x,y)$ of (54) having the properties mentioned after (35).

Introduce θ, ϕ as new independent variables. Then, since (54) implies that $(\lambda+i\mu)dx+dy = (\theta_y+i\phi_y)^{-1}(d\theta+id\phi)$, the relation (53) becomes

$$\int_J (U+iV)(\theta_y+i\phi_y)^{-1}(d\theta+id\phi) = \iint_E (\gamma^*U+\delta^*V)\,d\theta d\phi,$$

where γ^*, δ^* are continuous functions of (θ,ϕ). If new dependent variables U^*, V^* are defined by $U^*+iV^* = (U+iV)(\theta_y+i\phi_y)^{-1}$, there results the identity

$$\int_J (U^*+iV^*)(d\theta+id\phi) = \iint_E (\alpha^{**}U^*+\beta^{**}V^*)\,d\theta d\phi,$$

where α^{**}, β^{**} are continuous functions of (θ,ϕ).

This shows that Theorem (*) is applicable if u, v, x, y in Theorem (*) are replaced by U^*, V^*, θ, ϕ, respectively. If the conclusions of Theorem (*) are re-stated in terms of the original variables u, v, x, y, Theorem (**) follows.

10. The Hessian. By generalizing an idea of Cohn-Vossen ([3]; cf. the proofs given by H. Hopf and H. Samelson [9] and by Zhitomirsky [15]) in the problem of embedding of a binary Riemannian analytic metric of positive curvature, H. Lewy [10] has formulated and proved a general theorem on the solutions of non-linear differential equation $F=0$ of elliptic type, under the assumption that the function $F = F(x,\cdot\cdot\cdot,t)$ is analytic in its eight variables and the solutions are analytic in (x,y). *Loc. cit.*, the restriction of analyticity is essential indeed; the proof ([10], pp. 259-260) does not apply even if the solutions are of class C^∞. It is therefore of interest that the procedures applied above lead to a proof which extends Lewy's theorem to the

case in which F is of class C^2 only and the solutions involved are of class C^3 (Theorem (†) in Section **12** below).

To this end, the following counterpart of Theorem 1 will first be proved:

THEOREM 3*. *Besides the assumptions made in Theorem* 1* *on the coefficient functions* $a, \cdots, f$ *of* (29), *suppose that* d, e, f, *too, are of class* C^1 *on* (1). *Then, if* u *is any solution, of class* C^2, *of* (29) *satisfying* (4) *for some positive integer* n, *both limits*

$$(55)\qquad \lim_{\rho\to 0}(u_{xy}+iu_{xx})/(x+iy)^{n-1}, \qquad \lim_{\rho\to 0}(u_{yy}+iu_{xy})/(x+iy)^{n-1} \text{ exist}$$

(*for the same* $n \geqq 1$).

If the two limits (55) are denoted by $n(c_2+ic_1)$, $n(c_4+ic_3)$, where c_k is real, then, since $\rho=(x^2+y^2)^{\frac{1}{2}}$, it follows from (55) that, to an error $o(\rho^{1-n})$ as $\rho\to 0$, the function $\rho^{1-n}u_{xy}/n$ of (x, y) is both

$$c_2\cos(n-1)\theta - c_1\sin(n-1)\theta \text{ and } c_3\cos(n-1)\theta + c_4\sin(n-1)\theta,$$

where $x+iy=\rho e^{i\theta}$. Hence $c_2-c_3=0$ and $c_1+c_4=0$. Consequently, from (55),

$$\begin{aligned} &u_{xx}=n\rho^{n-1}(c_1\sin(n-1)\theta+c_2\sin(n-1)\theta)+o(\rho^{n-1}),\\ (55\text{ bis})\quad &u_{xy}=n\rho^{n-1}(c_2\cos(n-1)\theta-c_1\sin(n-1)\theta)+o(\rho^{n-1}),\\ &u_{yy}=n\rho^{n-1}(-c_1\cos(n-1)\theta-c_2\sin(n-1)\theta)+o(\rho^{n-1}).\end{aligned}$$

Clearly, (5′) follows from (55 bis) in the same way as (5″) did from (5′).

Note that the assumption in Theorem 3* is $n\geqq 1$ (whereas it is $n\geqq 0$ in Theorem 1*); so that $\operatorname{grad} u=0$ at $(x,y)=(0,0)$. This assumption cannot, in general, be replaced by

$$(56)\qquad u(x,y)=u(0,0)+p(0,0)x+q(0,0)y+o(\rho^n);$$

cf. the counterexamples given in connection with (8), Section **2**. Of course, it need not be assumed in Theorem 3* that $u(0,0)=0$ if $f\equiv 0$; similarly, (56) can replace (4) if $d\equiv e\equiv f\equiv 0$.

Since Theorem 2* shows that the limits in (55) cannot be 0 for *every* positive integer n when $u\not\equiv 0$, it follows that the "flat points" of the surface $u=u(x,y)$, that is, the points where the Hessian matrix of u is the zero matrix, cannot cluster at $(x,y)=(0,0)$ unless the surface is the plane $u\equiv 0$. In this direction, Theorem 3* contains a stronger statement:

Corollary 2*. *Under the assumptions of Theorem* 3*, *there exists an* $\epsilon > 0$ *such that*

$$u_{xx}u_{yy} - u_{xy}^2 < 0 \text{ for } 0 < x^2 + y^2 < \epsilon^2 \tag{57}$$

unless $u \equiv 0$.

In other words, the Gaussian curvature of the surface $u = u(x, y) \not\equiv 0$ is negative in a punctured vicinity of $(x, y) = (0, 0)$. Thus $(x, y) = (0, 0)$ is either a hyperbolic or an *isolated* parabolic point of the surface.

11. Proof of Theorem 3. It can be assumed that

$$\Delta = (ac - b^2)^{\frac{1}{2}} \equiv 1, \tag{58}$$

since (29) can be multiplied by $1/\Delta$.

If u is of class C^3, differentiation of (29) with respect to x, y and an application of Green's formula give the identities

$$\int_J (br + cs)\,dx - (ar + bs)\,dy = \int\int_E I_1 dxdy, \tag{59}$$

$$\int_J (bs + ct)\,dx - (as + bt)\,dy = \int\int_E I_2 dxdy, \tag{60}$$

where E is any subdomain E of (1) bounded by a piecewise smooth (C^1) Jordan curve J in (1), and

$$I_1 = (d - b_y)r + (e + b_x - c_y)s + c_x t + d_x p + e_x q + f_x u, \tag{61}$$

$$I_2 = a_y r + (d + b_y - a_x)s + (e - b_x)t + d_y p + e_y q + f_y u. \tag{62}$$

Actually, these integral relations can be proved under the assumptions of Theorem 3*, where it is only supposed that u is of class C^2.

To this end, note that (29) gives

$$-\int_J (ar + bs)dy = \int_J bsdy + \int_J ctdy + \int_J (dp + eq + fu)dy. \tag{63}$$

The last integral can be written as a double integral,

$$\int_J (dp + eq + fu)dy = \int\int_E (dr + es + fp + d_x p + e_x q + f_x u)dxdy. \tag{64}$$

The first integral on the right of (63) can be treated as follows:

$$\int_J bsdy = -\int_J brdx + \int_J b(rdx + sdy),$$

where

$$\int_J b(rdx + sdy) = \int\int_E (b_x s - b_y r)\,dxdy,$$

by the Lemma in [5], p. 761. Thus

$$\int_J bsdy = -\int_J brdx + \int\int_E (b_x s - b_y r)\,dxdy. \tag{65}$$

Similarly,

$$\int_J ctdy = -\int_J csdx + \int\int_E (c_x t - c_y s)\,dxdy. \tag{66}$$

But (59) follows from (63), (64), (65) and (66). The relation (60) is proved similarly.

Let (36) be a solution of (35) such that (36) has a non-vanishing Jacobian, maps (1) onto some domain D^* in a one-to-one manner and (40) holds. Introduce the abbreviations

$$U(\theta,\phi) \equiv u(x,y), \qquad P(\theta,\phi) \equiv p(x,y), \qquad Q(\theta,\phi) \equiv q(x,y). \tag{67}$$

Then (59), (60) can be writen as

$$\int_J P_\phi d\theta - P_\theta d\phi = \int\int_E I^*_1 d\theta d\phi, \tag{68}$$

$$\int_J Q_\phi d\theta - Q_\theta d\phi = \int\int_E I^*_2 d\theta d\phi, \tag{69}$$

where I^*_1, I^*_2 are linear forms in U, P, Q, P_θ, P_ϕ, Q_θ, Q_ϕ, with coefficients which are continuous functions of (θ, ϕ). Also

$$U(\theta,\phi) = \int_{(0,0)}^{(\theta,\phi)} (Px_\theta + Qy_\theta)\,d\theta + (Px_\phi + Qy_\phi)\,d\phi.$$

For a moment, let (θ, ϕ) be renamed (x, y). Then (14) holds for $w = w_1$ or $w = w_2$ and $W = W_1$ or $W = W_2$, respectively, where $w_1 = P_y + iP_x$, $W_1 = I^*_1$ and $w_2 = Q_y + iQ_x$, $W_2 = I^*_2$. Correspondingly, (17_k) holds for

$w = w_1$ or $w = w_2$ if (16_k) holds for $w = w_1$ or $w = w_2$, respectively. If $R > 0$ is sufficiently small, then there exists a constant $K = K_R > 0$ satisfying

$$| W_1 | + | W_2 | \leqq K(| U | + | P | + |Q| + | w_1 | + | w_2 |) \text{ for } | z | \leqq R.$$

Hence, if $| w_1 | + | w_2 |$ is denoted by ω, the inequality corresponding to (19_k) is

$$2\pi\omega(\zeta) | \zeta |^{-k} \leqq \int_{|z|=R} \omega | z^{-k}(z-\zeta)^{-1} | \cdot | dz |$$
$$+ K \iint_{|z|<R} (\omega + | P | + | Q | + | U |) | z^{-k}(z-\zeta)^{-1} | \, dxdy.$$

Since $| U | + | P | + | Q | = o(\rho^{1-n})$ as $\rho \to 0$, obvious modifications of the last part of the proof of Theorem 1 show that w_1/z^{n-1} and w_2/z^{n-1}, where $z = x + iy$, tend to limits as $z \to 0$.

In the original notation, this means that, as $(\theta, \phi) \to (0, 0)$,

$$\lim (P_\phi + iP_\theta)/(\theta + i\phi)^{n-1} \text{ and } \lim (Q_\phi + iQ_\theta)/(\theta + i\phi)^{n-1}$$

exist. It is seen from the proof of Theorem 2* that this completes the proof of (55), which is the statement of Theorem 3*.

Proof of Corollary 2*. Let $n(> 1)$ be the least positive integer for which not both limits (55) are 0; that is, the least positive integer for which (4) does not hold. Then not both constants c_1, c_2 are 0 in (55 bis) and (5′), (5″). Since (55 bis) implies that

$$u_{xx}u_{yy} - u_{xy}^2 = -\rho^{2n-2}(c_1^2 + c_2^2) + o(\rho^{2n-2}),$$

Corollary 2* follows.

12. Non-linear elliptic equations. It has been observed by Hadamard [4], pp. 352-354, that, in proving uniqueness theorems for partial differential equations

$$(70) \qquad F(x, y, u, p, q, r, s, t) = 0,$$

it is usually sufficient to consider the linear cases (29). For, if F in (70) is a function of class C^n, with $n \geqq 1$, on some eight-dimensional, convex domain $\mathfrak{D}$, and if $u = u_1(x, y)$ and $u = u_2(x, y)$ are solutions of (70) of class C^k, where $k \geqq 2$, on some (x, y)-domain D, then the difference $u = u_1 - u_2$ is a solution of a linear equation (29), in which the functions

$a, b, \cdots, f$ of (x, y) are of class C^m, where $m = \min(n-1, k-2)$. If in addition $4F_rF_t - F_s^2 \neq 0$ on $\mathfrak{D}$, then $\operatorname{sgn}(ac - b^2) = \operatorname{sgn}(4F_rF_t - F_s^2)$. Hence, Theorems 1*, 2*, 3* and their Corollaries imply the following theorem:

THEOREM (†). *Let $F(x, y, u, p, q, r, s, t)$ be a function of class C^2 on an eight-dimensional convex domain $\mathfrak{D}$ on which*

$$4F_rF_t - F_s^2 > 0. \tag{71}$$

On some connected (x, y)-domain D, let $u_1(x, y)$ and $u_2(x, y)$ be two solutions of (70), both of class C^3, and suppose that the function $u = u_1 - u_2$ of (x, y) satisfies

$$u = 0 \text{ at } (x, y) = (0, 0) \tag{72}$$

but $u \not\equiv 0$ (i. e., $u_1 \not\equiv u_2$) on D. Then (4) cannot hold for every n and, if $\epsilon > 0$ is small enough,

$$u_x^2 + u_y^2 \neq 0 \text{ for } 0 < x^2 + y^2 < \epsilon^2. \tag{73}$$

If, in addition to (72), it is assumed that

$$u_x^2 + u_y^2 = 0 \text{ at } (x, y) = (0, 0), \tag{74}$$

then ϵ can be chosen so small that, in addition to (73),

$$u_{xx}u_{yy} - u_{xy}^2 < 0 \text{ for } 0 < x^2 + y^2 < \epsilon^2. \tag{75}$$

In his short note [1], Carleman states that a weakened form of the conclusion (73), namely, $u^2 + u_x^2 + u_y^2 \neq 0$ for $0 < x^2 + y^2 < \epsilon^2$, follows from his result on linear elliptic systems of the form (9 bis), where α, β, γ, δ are continuous functions of (x, y). But this argument does not seem to be obvious. It is certainly valid if the linear partial differential equation (29) satisfied by the difference $u = u_1 - u_2$ is of the type (2), with $f \equiv 0$. For then (2) can be written as the system $q_x - p_y = 0$, $q_y + p_x = -eq - dp$. But if the coefficients a, b, c are only of class C^1, then (29) cannot in general be reduced to the form (2), but only to the integrated form (3) of (2); cf. the Remark following (39). After this difficulty, the case $f \not\equiv 0$ would still remain to be dealt with.

In this connection it is worth mentioning that *the assumption* (72) *is not needed for the conclusions* (73) *and* (75) *if F does not depend on u explicitly* and that *neither of the assumptions* (72), (74), *is needed for* (75) *if F does not contain* (u, p, q).

If $u = u_1(x, y)$ and $u = u_2(x, y)$ are analytic functions (i. e., expand-

able in power series) of (x, y) and if F, too, is an analytic function (of $(x, y, u, \cdots, t)$), then the conclusion (75) is due to Lewy [10], pp. 259-260.

13. Non-parabolic linear systems. In what follows, a uniqueness theorem on linear systems will be considered. It will be proved under assumptions of differentiability lighter than those imposed by Carleman [2]. This will depend on a suitable modification and, at the same time, simplification of Carleman's proof (some points omitted by Carleman [2], p. 8, will be proved in detail; cf. the motivations in (87)-(90) below). The refined form of Carleman's theorem is as follows:

THEOREM (§). *Let $A = A(x, y)$ and $B = B(x, y)$ be n by n matrix functions of class C^1 and of class C^0 (= continuous), respectively, on the closure of the (x, y)-domain*

$$D_R\colon\ x > 0, x^2 + y^2 < R^2, \tag{76}$$

and let the elementary divisors of A be simple. Let $u = u(x, y)$ be any vector function, with n components, which is of class C^1 and satisfies

$$u_x + Au_y = Bu \tag{77}$$

on (76), is continuous on the closure of (76) and, if $u(0, y)$ denotes the boundary value $u(+0, y)$, let

$$u(0, y) \equiv 0 \tag{78}$$

for $|y| < R$. Then there exists a positive $\epsilon(< R)$ such that $u(x, y) \equiv 0$ on D_ϵ.

The proof of Theorem (§) will use the fact that, by the assumptions made for A, the definition of the function A can be extended from the domain D_R in such a way that A becomes of class C^1 on a domain containing the closure of D_R. In the statement of his theorem, Carleman ([2], p. 1) requires that A be of class C^2 (instead of class C^1) and that B be of class C^0 on D_R (not on the closure of D_R). Carleman makes no assumption on A and B as $x \to +0$, but this must be an oversight; for he obviously uses the fact that B is bounded on (76) and, in his use of Green's formula, he implicitly imposes (unspecified) conditions on A as $x \to +0$.

REMARK. Corresponding to the Remark following Corollary 2 in Section 1, Theorem (§) remains true if the linear terms Bu on the right of (77) are modified by continuous non-linear terms satisfying conditions similar to those specified by the formula line in that Remark. This will be clear from the proof of Theorem (§).

14. Proof of Theorem (§). The assumptions made for A imply that there exists on the closure of D_R a non-singular matrix $T = T(x, y)$, of class C^1, for which $T^{-1}AT$ becomes a matrix with binary blocks of the form

$$(79) \qquad \begin{pmatrix} \lambda_j & -\mu_j \\ \mu_j & \lambda_j \end{pmatrix}, \ \mu_j \neq 0, \qquad j = 1, \cdots, m$$

along its main diagonal, where $0 \leqq m \leqq \frac{1}{2}n$, and with $n - 2m$ diagonal element λ_j, where $j = 2m + 1, \cdots, n$, while all other elements of $T^{-1}AT$ are zero; cf. [6], p. 856. Hence, if u is replaced by the vector Tu, it can be supposed that A itself has the normal form, just described. Thus, if $u = (u^1, \cdots, u^n)$ and $B = (b_{jk})$, the system (1) is of the form

$$(80_j) \qquad \begin{aligned} &u^j{}_x + \lambda_j u^j{}_y - \mu_j u^{j+1}{}_y = \sum_{k=1}^{n} b_{jk} u^k, \\ &\qquad\qquad\qquad\qquad \text{where } j = 1, 3, \cdots, 2m-1 \\ &u^{j+1}{}_x + \mu_j u^j{}_y + \lambda_j u^{j+1}{}_y = \sum_{k=1}^{n} b_{j+1\,k} u^k \end{aligned}$$

and

$$(81_j) \qquad u^j{}_x + \lambda_j u^j{}_y = \sum_{k=1}^{n} b_{jk} u^k, \qquad \text{where } j = 2m + 1, \cdots, n.$$

The functions $\mu_1, \cdots, \mu_m;\ \lambda_1, \cdots, \lambda_m;\ \lambda_{2m+1}, \cdots, \lambda_n$ are of class C^1 on the closure of D_R and can, therefore, be extended so as to become of class C^1 on (1).

It can be supposed that $m > 0$, for otherwise (77) is a hyperbolic system for which Theorem (§) seems to be well-known, even though B is just continuous (cf. a statement of Carleman [2], p. 2; the assumptions made by Holmgren [8] are heavier; see also [6], pp. 855-864, where on p. 864, the first reference to Holmgren is erroneous and must be replaced by [8] in the bibliography of the present paper). There corresponds to the pair $(\lambda, \mu) = (\lambda_j, \mu_j)$ the Cauchy-Riemann-Beltrami system (54), which has on (1) a solution $\theta = \theta_j(x, y)$, $\phi = \phi_j(x, y)$, of class C^1 and of non-vanishing Jacobian, satisfying

$$(82) \qquad \theta_j(0, 0) = \phi_j(0, 0) = 0$$

and mapping (1) in a one-to-one manner on a (θ_j, ϕ_j)-domain D^j.

Let $c > 0$, $\tau > 1$ be such that the portion $x > 0$ of the circle $|z + c| \leqq \tau c$, where $z = x + iy$, is contained in D_R. Let this portion of D_R be denoted by $D(c)$, and let $\Gamma(c)$ denote the arc $x > 0$ of the circle $|z + c| = \tau c$; finally, let the images of $D(c)$, $\Gamma(c)$ under the transformation $(x, y) \to (\theta_j, \phi_j)$ be

called $D^j(c)$, $\Gamma^j(c)$, respectively. The pair of equations (80_j) on $D(c)$ is transformed into equations of the form

$$u^j{}_\theta - u^{j+1}{}_\phi = \sum_{k=1}^n b_k u^k, \qquad u^j{}_\phi + u^{j+1}{}_\theta = \sum_{k=1}^n b^*{}_k u^k, \qquad (\theta, \phi) = (\theta_j, \phi_j),$$

on $D^j(c)$, where b_k, $b^*{}_k$ and u are continuous functions of (θ, ϕ) on the closure of $D^j(c)$. (Incidentally, the coefficients b_k, $b^*{}_k$ depend in a very simple manner on b_{jk}, $b_{j+1\,k}$, on the partial derivatives θ_x, θ_y, ϕ_x, ϕ_y and on the Jacobian $\partial(\theta, \phi)/\partial(x, y) \neq 0$.) It follows that there exists a constant K (independent of j and c, for small $c > 0$) for which

$$(83) \qquad \iint_D | w^j(\zeta)/(\zeta - \zeta_0)^l | \, d\theta d\phi$$

$$\leqq Kc \int_\Gamma | w^j(\zeta)/(\zeta - \zeta_0)^l | \cdot | d\zeta | + Kc \iint_D \sum_{k=1}^n | u^k/(\zeta - \zeta_0)^l | \, d\theta d\phi,$$

where l is a positive integer, $w^j = u^{j+1} + iu^j$, $D = D^j(c)$, $\Gamma = \Gamma^j(c)$, $\zeta = \theta + i\phi = \theta_j + i\phi_j$, and ζ_0 is the $(\theta_j + i\phi_j)$-image of $x + iy = -c + i0$; cf. the derivation of (24) above.

Change the integration variables from $(\theta, \phi) \equiv (\theta_j, \phi_j)$ to (x, y) in (83). The function $| \partial(\theta, \phi)/\partial(x, y) |$ is bounded from above and from below by positive constants (independent of c, for small $c > 0$, on $D(c)$); also, the ratio $| \zeta - \zeta_0 |/| z + c |$ is bounded from above and from below by positive constants; finally, $| d\zeta |$ is majorized by Const. $| dz |$, since $(d\theta + id\phi) = (\theta_y + i\phi_y) dz$. Hence, (83) implies that, if the constant K is large enough,

$$(84_j) \qquad \iint_D \{(| u^j | + | u^{j+1} |)/| z + c |^l\} dxdy$$

$$\leqq Kc \int_\Gamma \{(| u^j | + | u^{j+1} |)/| z + c |^l\} | dz | + Kc \sum_{k=1}^n \iint_D | u^k/(z + c)^l | \, dxdy,$$

where $D = D(c)$ and $\Gamma = \Gamma(c)$. Addition of $(84_1), (84_3), \cdots, (84_{2m-1})$ gives

$$(85) \qquad (1 - Kc) \sum_{j=1}^{2m} \iint_D | u^j/(z + c)^l | \, dxdy$$

$$\leqq Kc \sum_{j=1}^{2m} \int_\Gamma | u^j/(z + c)^l | \cdot | dz | + Kc \sum_{k=2m+1}^{n} \iint_D | u^k/(z + c)^l | \, dxdy,$$

where $D = D(c)$ and $\Gamma = \Gamma(c)$.

In order to appraise the last integral, let $y = y_j = y(x; \xi, \eta)$, where $y(\xi; \xi, \eta) = \eta$, denote the (unique) solution of the differential equation

$$(86_j) \qquad dy/dx = \lambda_j(x, y) \qquad (j = 2m + 1, \cdots, n)$$

of the "characteristics" of (81_j). Then, according to (81_j) and the assumption $u^j(0, y) \equiv 0$,

$$(87) \qquad u^j(x, y) = \sum_{k=1}^{n} \int_0^x b_{jk}(t, y(t; x, y)) u^k(t, y(t; x, y)) dt.$$

If c and τ — (> 0) are small, the absolute value of the slope of the arc $\Gamma(c)$ exceeds any given constant, hence, in particular, the function $|\lambda_j(x, y)|$. Hence (86_j) implies that if the point (x, y) is in $D(c)$, then the arc $(t, y(t; x, y))$, where $0 < t < x$, is in $D(c)$; in fact, $|t + iy(t; x, y) + c|$ is an increasing function of t. Consequently, the constant K can be chosen so large that

$$(88) \qquad |u^j(x, y)/(z + c)^l|$$
$$\leqq K \sum_{k=1}^{n} \int_0^x |u^k(t, y(t; x, y))/(t + c + iy(t; x, y))^l| \, dt$$

for all (x, y) in $D(c)$. Integration of this inequality with respect to $dxdy$ over $D = D(c)$ gives

$$(89) \qquad \iint_D |u^j/(z + c)^l| \, dxdy$$
$$\leqq K \sum_{k=1}^{n} \iint_D \int_0^x |u^k(t; y(t; x, y))/(t + c + iy(t; x, y))^l| \, dtdxdy.$$

Let the integration variables (x, y) in the triple integral on the right of (89) be replaced by the variables (ρ, θ), where $z + c = \rho e^{i\theta}$. Then, if $\theta = \theta(\rho) = \arccos c/\rho$ and $0 < \rho(\theta) < \frac{1}{2}\pi$, the triple integral becomes of the type

$$\int_c^{\tau c} \rho d\rho \{ \int_{-\theta(\rho)}^{\theta(\rho)} \int_0^{\rho \cos\theta} h(t; y(t; \rho\cos\theta, \rho\sin\theta)) dt d\theta \}$$

where $h = h(x, y) = |u^k(x, y)/(z + c)^l|$. If the integration variables (t, θ) in the inner double integral are replaced by (x, y), where $x = t$, $y = y(t; \rho\cos\theta, \rho\sin\theta)$, then the domain of integration for (x, y) is $E: x > 0$, $|z + c| < \rho$; cf. the remark following (87). Furthermore, the Jacobian

$$\partial(x,y)/\partial(t,\theta) = \rho \sin\theta\, \partial y(t;\xi,\eta)/\partial\xi + \rho\cos\theta\, \partial y(t;\xi,\eta)/\partial\eta,$$

where $(\xi,\eta) = (\rho\cos\theta, \rho\sin\theta)$, lies between two positive constant multiples of ρ, since $|\sin\theta|$ is small and $\partial y(0;0,\eta)/\partial\eta \equiv 1$. Hence the last triple integral is majorized by

$$K\int_c^{Tc} d\rho\{\int\int_E h(x,y)dxdy\},$$

where K is a constant. Consequently, (89) leads to the inequality

$$\int\int_D |u^j/(z+c)^l|\,dxdy \leqq Kc\sum_{k=1}^n \int\int_D |u^k/(z+c)^l|\,dxdy,$$

where $D = D(c)$, $j = 2m+1,\cdots,n$, and $l = 1,2,\cdots$.

Addition of these inequalities gives

$$(90)\qquad (1-Kc)\sum_{j=2m+1}^n \int\int_D |u^j/(z+c)^l|\,dxdy$$
$$\leqq Kc\sum_{k=1}^{2m}\int\int_D |u^k/(z+c)^l|\,dxdy,$$

where $D = D(c)$ and $l = 1,2,\cdots$. If $c>0$ is so small that $1-Kc>\frac{1}{2}$, then the last inequality, when combined with (85), leads to

$$(91)\qquad (1-3Kc)\sum_{j=1}^{2m}\int\int_D |u^j/(z+c)^l|\,dxdy$$
$$\leqq Kc\sum_{k=1}^{2m}\int_\Gamma |u^k/(z+c)^l|\cdot|dz|.$$

As at the end of the proof of Theorem 2, it follows that (91) cannot hold for large l unless $u^j \equiv 0$ on $D = D(c)$ for $j = 1,2,\cdots,2m$, in which case (90) shows that $u^j \equiv 0$ on $D = D(c)$ for $j = 2m+1,\cdots,n$. This proves Theorem (§).

REMARK. If (77) is an elliptic system, that is, if no eigenvalue of $A(x,y)$ is real, then the hypothesis that u vanishes on a segment of the y-axis can be replaced by the assumption that $u=0$ holds on a sequence of point of (76) clustering at an interior point of the domain of definition of u. (Actually, if, in the wording of Theorem (§) above, u is defined to be iden-

tically 0 for $x < 0, x^2 + y^2 < R^2$, then u is a continuous solution of an integrated form of (77) on (1).) On the other hand, if A has at least one real characteristic number, then the assumption $u \equiv 0$ for $x = 0$ cannot be replaced by the assumption that $u = 0$ holds on a set of points clustering at a point of D_R. In order to see this, it is sufficient to choose (77) to be a scalar equation, say $u_x = 0$.

15. Non-parabolic non-linear systems. The linearizing device of Hadamard, used above (Section **12**) in the reduction of Theorem (†) to Theorems 1*, 2*, 3*, also shows that Theorem (§) has the following consequence:

THEOREM (∇). *Let $F = F(x, y, u, w)$, where F, u, w are vectors with n components, be a function of class C^2 on a $(2n+2)$-dimensional domain $\mathfrak{D}$ which contains the product set of the (x, y)-domain D_R defined by (76), and of a $2n$-dimension convex (u, w)-domain, and suppose that the elementary divisors of the Jacobian matrix* $\operatorname{grad}_w F$ *are simple at every point of $\mathfrak{D}$. On D_R, let $u = u_1(x, y)$ and $u = u_2(x, y)$ be two solutions, of class C^2, of the system of n differential equations*

$$u_x = F(x, y, u, u_y). \tag{92}$$

Then, if $u_1(+0, y)$, $u_2(+0, y)$ exist uniformly and are identical for $-R < y < R$, there exists an $\epsilon > 0$ such that $u_1(x, y) \equiv u_2(x, y)$ on D_ϵ.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] T. Carleman, "Sur les systèmes linéaires aux dérivées partielles du premier ordre à deux variables," *Comptes Rendus*, vol. 197 (1933), pp. 471-474.

[2] ———, "Sur un problème d'unicité pour les systèmes d'équations aux dérivées partielles à deux variables indépendantes," *Arkiv för Matematik, Astronomi och Fysik*, vol. 26B (1938), no. 17.

[3] St. Cohn-Vossen, "Zwei Sätze über die Starrheit der Eiflächen," *Göttinger Nachrichten*, 1927, pp. 125-134.

[4] J. Hadamard, *Leçon sur la propagation des ondes*, Paris, 1903.

[5] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[6] ——— and A. Wintner, "On hyperbolic partial differential equations," *ibid.*, vol. 74 (1952), pp. 834-864.

[7] ——— and A. Wintner, "On the existence of Riemannian manifolds which cannot carry non-constant analytic or harmonic functions in the small," *ibid.*, vol. 75 (1953), pp. 260-276.

[8] E. Holmgren, "Sur les systèmes linéaires aux dérivées partielles du premier ordre à deux variables indépendentes à caractéristiques réelles et distinctes," *Arkiv för Matematik, Astronomi och Fysik*, vol. 5 (1909), no. 1.

[9] H. Hopf and H. Samelson, "Zum Beweis des Kongruenzsatzes für Eiflächen," *Mathematische Zeitschrift*, vol. 43 (1938), pp. 749-766.

[10] H. Lewy, "On differential geometry in the large, I," *Transactions of the American Mathematical Society*, vol. 43 (1938), pp. 258-270.

[11] L. Lichtenstein, "Zur Theorie der konformen Abbildung. Konforme Abbildung nicht-analytischer, singularitätenfreier Flächenstücke auf ebene Gebiete," *Bulletin International de l'Académie des Sciences de Cracovie*, ser. A, 1916, pp. 192-217.

[12] ———, "Neuere Entwicklung der Theorie partieller Differentialgleichungen zweiter Ordnung vom elliptische Typus," *Encyklopädie der mathematischen Wissenschaften*, vol. II C (1924), no. 12, pp. 1279-1334.

[13] H. Petrini, "Les dérivées premieres et secondes du potentiel logarithmique," *Journal de Mathématiques*, ser. 6, vol. 5 (1909), pp. 127-223.

[14] A. Wintner, "On the Hölder restriction in the theory of partial differential equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 731-738.

[15] O. K. Zhitomirsky, "Sur la non-flexibilité des ovaloïdes," *Comptes Rendus (Doklady) URSS*, vol. 25 (1939), pp. 347-349.

ON PIECES OF CONVEX SURFACES.*

By Philip Hartman and Aurel Wintner.

1. Let the vector function $N = N(u, v)$ be either of the continuous unit normals of a sufficiently small piece S of a surface which is of class C^1 and is given in the parametric form $X = X(u, v)$, where $X = (x, y, z)$ is a function of class C^1 on a simply connected (u, v)-domain and the vector product (X_u, X_v) does not vanish. If $X = X(u, v)$ is of class C^2 and the Gaussian curvature $K = K(u, v)$ is not zero, then the correspondence between points $X(u, v)$ of S and points $N(u, v)$ of the unit sphere, $|N| = 1$, is one-to-one in a neighborhood of every point (u^0, v^0). Such a correspondence can exist when the assumption $K \neq 0$ is weakened to $K \geqq 0$ or $K \leqq 0$, but cannot exist if K assumes both positive and negative values in every neighborhood of the point (u^0, v^0) under consideration.

In order to verify this negative statement, note that if

$$p = p(x, y), \qquad q = q(x, y) \tag{1}$$

is a one-to-one continuous mapping of a simply connected (x, y)-domain D onto a (p, q)-domain, then the orientation of the image in the (p, q)-plane of a positively oriented Jordan curve J in D is independent of J. If in addition the functions (1) are of class C^1 and if (1) has a non-vanishing Jacobian, then the orientation of the image in the (p, q)-plane is positive or negative according as the Jacobian $\partial(p, q)/\partial(x, y)$ is positive or negative. Hence, if the functions (1) are of class C^1, then (1) cannot be a one-to-one, continuous mapping of one simply connected domain onto another when $\partial(p, q)/\partial(x, y)$ assumes positive and negative values.

If a surface S of class C^1 has a parametrization $X = (x, y, z(x, y))$, where $z(x, y)$ is of class C^1 on a (small) simply connected domain D, then the correspondence between points X of S and points N of the unit sphere is essentially given by (1), where

$$p = z_x(x, y), \qquad q = z_y(x, y). \tag{2}$$

In fact, the direction cosines (λ, μ, ν) of $\pm N$ are determined by

$$\lambda = p/(1 + p^2 + q^2)^{\frac{1}{2}}, \qquad \mu = q/(1 + p^2 + q^2)^{\frac{1}{2}} \tag{3}$$

* Received September 25, 1952.

and $\lambda^2 + \mu^2 + \nu^2 = 1$; so that the mapping $X \to N$ is continuous and one-to-one if and only if (2) is. When S (hence z) is of class C^2, the Jacobian of (2) is $z_{xx}z_{yy} - z_{xy}^2$, which is $(1 + p^2 + q^2)^2$ times the Gaussian curvature $K = K(x, y)$. Since this implies that $\operatorname{sgn} \partial(p, q)/\partial(x, y) = \operatorname{sgn} K$, the negative statement, made before (1), follows.

2. There will be considered an analogue of the statement in the first paragraph concerning the correspondence $X \to N$ when $K \neq 0$. In this analogue, it will only be assumed that S is of class C^1; in particular, S need not possess a Gaussian curavture K. The assumption $K > 0$ or $K < 0$ will be replaced by the hypotheses that S is strictly convex (which implies that $K \geqq 0$ if S is of class C^2).

A continuous function

$$z = z(x, y), \tag{4}$$

defined on a convex neighborhood D of $(x, y) = (0, 0)$, is called *strictly convex* if

$$z(\tfrac{1}{2}\{x_1 + x_2\}, \tfrac{1}{2}\{y_1 + y_2\}) < \tfrac{1}{2}\{z(x_1, y_1) + z(x_2, y_2)\} \tag{5}$$

holds for every pair of distinct points (x_1, y_1), (x_2, y_2) of D. The function (4) is called convex if the $<$ in (5) is weakened to $\leqq$. A surface S will be called [strictly] convex if, after a suitable rotation of the (x, y, z)-space, it has a parametrization $X = X(x, y, z(x, y))$ in which (4) is [strictly] convex.

Let S be a surface of class C^1. By a spherical C^0-parametrization $X = X(\lambda, \mu)$ of S is meant a continuous, one-to-one mapping of a plane (λ, μ)-domain onto the surface S in such a way that λ, μ are two of the three direction cosines of a unit normal N at the point $X(\lambda, \mu)$ of S; for example, $N = (\lambda, \mu, \nu)$; cf. [2], pp. 305-306.

(I) *Let S be a strictly convex surface of class C^1. Then S has a spherical C^0-parametrization $X = X(\lambda, \mu)$ and, with respect to this parametrization, the supporting function*

$$H = X \cdot N \tag{6}$$

is of class C^1 (even though $X(\lambda, \mu)$ is just of class C^0).

It can be supposed that S has a parametrization $X = (x, y, z(x, y))$ in which (4) is of class C^1 and strictly convex in a convex neighborhood D of $(x, y) = (0, 0)$. The supporting function (6), in terms of (x, y), is given by

$$H = h/(1 + p^2 + q^2)^{\frac{1}{2}}, \tag{7}$$

where

$$h = xp + yq - z, \tag{8}$$

and p, q are defined by (2). In view of the remark made above on (2)-(3), the assertion (I) will be proved if it is shown that the transformation $(x, y) \rightarrow (p, q)$, defined by (2), is a continuous, one-to-one mapping of D onto a (p, q)-domain, and that h is of class C^1 as a function of (p, q). Hence (I) is contained in the following assertion:

(II) *Let $z = z(x, y)$ be a strictly convex function of class C^1 on a convex domain D. Then* (2) *is a continuous, one-to-one mapping of D onto a (p, q)-domain $\mathfrak{D}$, and* (8), *as a function of (p, q), is of class C^1 and strictly convex on every convex subdomain of $\mathfrak{D}$; finally, the inverse of* (2) *is*

$$x = h_p(p, q), \qquad y = h_q(p, q). \tag{9}$$

Thus the Legendre transformation $(x, y; z) \rightarrow (p, q; h)$, *along with the C^1-assumption,* is involutory.

The assertion of (II) has been proved by H. Lewy [3], pp. 365-366, under more restrictive conditions on z; namely, under the assumptions that z is of the form $z = \frac{1}{2}(x^2 + y^2) + z_1(x, y)$, where $z_1(x, y)$ is convex and of class C^1. The latter assumptions mean that, if z is smooth, the Gaussian curvature of the surface (4) satisfies $K > 0$, whereas the assumption of strict convexity only assures that $K \geqq 0$. It should also be noted that the involutory character, pointed out after (9), is lost if the functions z considered are those satisfying the conditions of Lewy's theorem (rather than the condition of strictly convex C^1-functions).

REMARK. *Theorems* (I), (II), *as well as* (IV) *below, have analogues for the case in which the surface S is replaced by a plane curve Γ.* Actually, these Γ-variants are corollaries of the S-theorems, since they result when S is a surface obtained by revolving Γ about a suitable axis.

3. The assertions of (I), (II) cannot be improved:

(III) *If the surface S: $z = z(x, y)$ is analytic* (*or even a polynomial*) *and strictly convex, then its supporting function H need not be of class C^2 in terms of spherical parameters; that is,* (8) *need not be of class C^2 as a function of (p, q).*

On the other hand, if S: $z = z(x, y)$ is of class C^n, where $n \geqq 2$, with a non-vanishing curvature K, then the supporting function is of class C^n as a function of (p, q); cf. [4], Appendix.

The truth of (III) is shown by the example $z(x,y) = (x^4 + y^4)/4$. In fact, an easy calculation shows that the corresponding function $h = h(p,q)$ is $3/4$ times $p^{4/3} + q^{4/3}$ and that the supporting function $H(p,q)$ is $3/4$ times $(p^{4/3} + q^{4/3})/(1 + p^2 + q^2)^{\frac{1}{2}}$.

It is worth observing that, in this example, (2) maps the convex domain D: $x^2 + y^2 < 1$ onto the non-convex domain $\mathfrak{D}: p^{2/3} + q^{2/3} < 1$.

4. *Proof of* (II). The assumptions made on $z(x,y)$ imply that if (x,y) is any point of D and θ is arbitrary, then $z(x + t\cos\theta, y + t\sin\theta)$ has a continuous derivative, $z_x\cos\theta + z_y\sin\theta$, with respect to t and that this derivative is strictly increasing with t. Hence, if (x',y'), (x'',y'') are two distinct points of D, then

$$z_x(x'',y'')\cos\theta + z_y(x'',y'')\sin\theta > z_x(x',y')\cos\theta + z_y(x',y')\sin\theta,$$

where $(\cos\theta, \sin\theta)$ denotes the unit vector in the direction of $(x'' - x', y'' - y')$. Thus, if (p',q'), (p'',q'') are points corresponding to (x',y'), (x'',y'') by virtue of (2), then

$$\Delta p\Delta x + \Delta q\Delta y > 0, \tag{10}$$

where $\Delta x = x'' - x'$, $\Delta p = p'' - p'$, etc. As is well known, (10) is necessary and sufficient that (4) be strictly convex; cf., e.g., [1], pp. 74-75 and p. 79.

Clearly, (10) implies the one-to-one character of the transformation (2).

Consider (8) as a function of (p,q) by virtue of the inverse of (2). Let (p_1,q), (p_2,q) be two points of $\mathfrak{D}$ and (x_1,y_1), (x_2,y_2) the corresponding points of D. Then the difference $\Delta h = h(p_2,q) - h(p_1,q)$ can be written as

$$\Delta h = x_1\Delta p + p_2\Delta x + q\Delta y - \Delta z,$$

where $\Delta p = p_2 - p_1$, $\Delta x = x_2 - x_1$, etc. The mean value theorem of differential calculus shows that $\Delta z = p^*\Delta x + q^*\Delta y$, where p^*, q^* are the partial derivatives of z at a point (x^*,y^*) on the segment joining (x_1,y_1) and (x_2,y_2). Hence

$$\Delta h = x_1\Delta p + (p_2 - p^*)\Delta x + (q - q^*)\Delta y.$$

Since the vector $(\Delta x, \Delta y)$ is a *positive* multiple of $(x_2 - x^*, y_2 - y^*)$, it follows from the case $(x',y') = (x^*,y^*)$, $(x'',y'') = (x_2,y_2)$ of (10) that $\Delta h > x_1\Delta p$. If Δh is written in the form

$$\Delta h = x_2\Delta p + p_1\Delta x + q\Delta y - \Delta z = x_2\Delta p - [(p^* - p_1)\Delta x + (q^* - q)\Delta y],$$

it follows that $\Delta h < x_2\Delta p$. Consequently, h has a partial derivative h_p with respect to p, and $h_p = x$. Similarly, h_q exists and is y. Since the inverse

$x = x(p, q)$, $y = y(p, q)$ of (2) is continuous, it follows that the partial derivatives of h are continuous.

In order to complete the proof of (II), it remains to ascertain that h is strictly convex. But this follows from (9) and (10); cf. [1], pp. 74-75 and p. 79.

REMARK. It remains undecided whether or not (II), hence a modified version of (I), remains correct if the assumption that S is strictly convex is replaced by the assumption that (2) is a one-to-one, continuous mapping of a simply connected (x, y)-domain D onto a (p, q)-domain $\mathfrak{D}$. Under this assumption, when D is suitably chosen, the set of points $z_x(x, y) = \text{const.}$ is empty or a Jordan arc and the family of these Jordan arcs covers D in a *schlicht* manner. A corresponding remark applies to the family $z_y(x, y) = \text{Const.}$ If it is further assumed that the arcs $z_x(x, y) = \text{const.}$, $z_y(x, y) = \text{Const.}$ are (locally) rectifiable, then those assertions of (II) which do not refer to convexity remain correct. For then, if J is the boundary of any rectangle $c_1 \leqq p \leqq c_2$, $C_1 \leqq q \leqq C_2$ in $\mathfrak{D}$, the image of J in D is rectifiable. If $x = x(p, q)$, $y = y(p, q)$ is the inverse of (2), then an integration by parts shows that

$$\text{(8 bis)} \qquad \int_J x\,dp + y\,dq = 0.$$

Hence there exists on $\mathfrak{D}$ a function $h = h(p, q)$ of class C^1 satisfying (9). The usual formula for obtaining h and another integration by parts show that such a function h is supplied by (8).

5. It was shown in [4], Theorem (iii), p. 368, that if S: $X = X(u, v)$ is a surface of class C^n, where $n \geqq 2$, and if c is a constant distinct from a pair of exceptional values c_1, c_2, then

$$\text{(11)} \qquad Y = Y(u, v) = X + cN$$

is a C^{n-1}-parametrization of a surface Σ which is of class C^n. The surface $\Sigma = \Sigma_c$ is called a parallel surface of S. It was also pointed out in [4], pp. 369-370, that if S is of class C^1, then the locus of the end points of (11) need not be a surface of class C^1 (for any choice of the constant c). In what follows, there will be proved an analogue of the result of Theorem (iii) of [4], an analogue in which $n \geqq 2$ is replaced by $n = 1$, but S is assumed to be convex.

(IV) *Let $z = z(x, y)$ be convex and of class C^1 on a convex neighbor-*

hood D of $(x,y)=(0,0)$ and let S denote the surface $X=(x,y,z(x,y))$: Let the constant c have the property that

$$(12)\qquad \alpha = x - cz_x(1+z_x^2+z_y^2)^{-\frac{1}{2}}, \qquad \beta = y - cz_x(1+z_x^2+z_y^2)^{-\frac{1}{2}}$$

is a one-to-one, continuous mapping of a sufficiently small D onto an (α,β)-domain $\mathfrak{D}$. (For instance, let $c<0$). Then the locus of end points of $Y=(\alpha,\beta,\gamma)$, where

$$(13)\qquad \gamma = z + c(1+z_x^2+z_y^2)^{-\frac{1}{2}}$$

is a surface Σ of class C^1, with $Y=(\alpha,\beta,\gamma(\alpha,\beta))$ as a C^1-parametrization. (If $c<0$, then $\gamma=\gamma(\alpha,\beta)$ is a convex function of (α,β) on every convex subdomain of $\mathfrak{D}$.)

Clearly, (11) is equivalent to (12)-(13).

It will first be verified that if $c<0$, then (12) is a one-to-one, continuous mapping. To this end, let (12) be written as $\alpha = x - c\lambda$, $\beta = y - c\mu$, where λ, μ are defined by (2)-(3). It is readily verified from (3) that

$$(14)\qquad \Delta\lambda\Delta p + \Delta\mu\Delta q \gtreqqless 0 \text{ according as } (\Delta p)^2 + (\Delta q)^2 \gtreqqless 0,$$

where the notation is similar to that in (10); in fact, (14) is a particular case of (10), where z is the strictly convex function $z=(1+x^2+y^2)^{\frac{1}{2}}$ and x, y, p, q are denoted by p, q, λ, μ, respectively. The convexity (rather than strict convexity) of z in (IV) implies that

$$(15)\qquad \Delta p\Delta x + \Delta q\Delta y \geqq 0.$$

Since

$$(16)\qquad \Delta\alpha\Delta p + \Delta\beta\Delta q = (\Delta x\Delta p + \Delta y\Delta q) - c(\Delta\lambda\Delta p + \Delta\mu\Delta q),$$

it follows that $\Delta\alpha\Delta p + \Delta\beta\Delta q > 0$ if $(\Delta p)^2 + (\Delta q)^2 > 0$ and $c<0$. On the other hand, $\Delta\alpha\Delta x + \Delta\beta\Delta y = (\Delta x)^2 + (\Delta y)^2$ if $(\Delta p)^2 + (\Delta q)^2 = 0$, by (12). Consequently, if $c<0$, then $(\Delta\alpha)^2 + (\Delta\beta)^2 > 0$ whenever $(\Delta x)^2 + (\Delta y)^2 > 0$; so that the one-to-one character of the continuous mapping (12) follows.

It should be remarked that (12) need not be a one-to-one transformation (of some small D) for any $c>0$, even though z is strictly convex. This is shown by an example of the type $z=g(x)+y^2$, where dg/dx is a continuous monotone function such that the sets of points where $d^2g/dx^2=0$ and where $d^2g/dx^2=\infty$, respectively, cluster at $x=0$.

As in the cases (I) and (II), it cannot be asserted that Σ is of class C^2, not even if additional smoothness conditions are imposed on S (and S is assumed to be strictly convex):

(V) *If* $z = z(x, y)$ *in* (IV) *is a polynomial which is strictly convex in the domain under consideration, and if* c *satisfies the conditions of* (IV), *then the parallel surface* $\Sigma = \Sigma_c$ *need not be of class* C^2.

In (IV), the constant c must, of course, be one of the exceptional values excluded in Theorem (iii) of [4], p. 368.

6. *Proof of* (IV). Since (12) is a one-to-one, continuous mapping of D onto $\mathfrak{D}$, the function (13) can be considered as a function of (α, β) on $\mathfrak{D}$ by virtue of the inverse of (12). Thus the set of points Σ which is claimed to be a surface of class C^1 has the parametrization $Y = (\alpha, \beta, \gamma(\alpha, \beta))$. Clearly, Σ is a surface of class C^1 if and only if $\gamma(\alpha, \beta)$ is a function of class C^1. It will be shown that γ is a function of class C^1 and that

$$(17) \qquad \gamma_\alpha = z_x, \qquad \gamma_\beta = z_y.$$

To this end, let (α_1, β), (α_2, β) be two points of $\mathfrak{D}$ and let (x_1, y_1), (x_2, y_2) be the corresponding points of D. Then the difference $\Delta\gamma = \gamma(\alpha_2, \beta) - \gamma(\alpha_1, \beta)$ can be written as

$$(18) \qquad \Delta\gamma = \Delta z + c\Delta\nu, \quad \text{where} \quad \nu = (1 - \lambda^2 - \mu^2)^{\frac{1}{2}};$$

cf. (2), (3). Also

$$(19) \qquad \Delta\alpha = \Delta x - c\Delta\lambda, \qquad \Delta\beta = 0 = \Delta y - c\Delta\mu,$$

by (12). In (18), Δz can be expressed as $p^*\Delta x + q^*\Delta y$, where p^*, q^* are the partial derivatives of z at a point (x^*, y^*) of the segment joining (x_1, y_1) and (x_2, y_2). Hence

$$\Delta\gamma = p_1\Delta x + q_1\Delta y + (p^* - p_1)\Delta x + (q^* - q_1)\Delta y + c\Delta\nu,$$

where (p_1, q_1), (p_2, q_2) correspond to (x_1, y_1), (x_2, y_2) by virtue of (2). If Δx, Δy are substituted from (19) into the last formula, there results

$$(20) \quad \Delta\gamma = p_1\Delta\alpha + [(p^* - p_1)\Delta x + (q^* - q_1)\Delta y] + c\{p_1\Delta\lambda + q_1\Delta\mu + \Delta\nu\}.$$

Similarly,

$$(21) \quad \Delta\gamma = p_2\Delta\alpha + [(p^* - p_2)\Delta x + (q^* - q_2)\Delta y] + c\{p_2\Delta\lambda + q_2\Delta\mu + \Delta\nu\}.$$

By virtue of the convexity of z, the term $[\cdot\ \cdot\ \cdot]$ in (20) is non-negative when the term $[\cdot\ \cdot\ \cdot]$ in (21) is non-positive, and conversely. Similarly, the terms $\{\cdot\ \cdot\ \cdot\}$ in (20) and (21) cannot be of the same sign, since their sum, with the respective positive weight factors $(1 + p_1^2 + q_1^2)^{-\frac{1}{2}}$, $(1 + p_2^2 + q_2^2)^{-\frac{1}{2}}$,

is the scalar product of the two vectors $(\lambda_2 \pm \lambda_1, \mu_2 \pm \mu_1, \nu_2 \pm \nu_1)$, which is 0, since both $(\lambda_i, \mu_i, \nu_i)$ are of length 1. Hence (20) and (21) show that

$$\min(p_1\Delta\alpha, p_2\Delta\alpha) \leqq \Delta\gamma \leqq \max(p_1\Delta\alpha, p_2\Delta\alpha).$$

This proves the first relation in (17), and the second is proved in the same way. Since z_x, z_y are continuous functions of (x, y) and since the inverse $x = x(\alpha, \beta)$, $y = y(\alpha, \beta)$ of (12) is continuous, it follows that γ is of class C^1.

It is clear from (14), (15), (16) and (17) that $\gamma = \gamma(\alpha, \beta)$ is a convex function (on convex subdomains of $\mathfrak{D}$) when $c < 0$. This completes the proof of (IV).

7. *Proof of* (V). For small $|x|$, consider the arc

$$\Gamma\colon z = \tfrac{1}{2}(x^2 + \tfrac{1}{2}x^4) \tag{22}$$

in the (x, z)-plane. A surface S proving the assertion (V) will be obtained by revolving (22) about the z-axis. Clearly, the parallel surfaces of S will be obtained by revolving the parallel curves

$$\alpha = x - c(x + x^3)(1 + (x + x^3)^2)^{-\frac{1}{2}}, \qquad \gamma = z + c(1 + (x + x^3)^2)^{-\frac{1}{2}} \tag{23}$$

of (22) in the (α, γ)-plane about the γ-axis. Thus, in order to discuss the parallel surfaces of S, it is sufficient to consider the parallel curves (23) of (22).

Expansion of $(1 + (x + x^3)^2)^{-\frac{1}{2}}$ into a power series $1 - \frac{1}{2}x^2 + \cdots$ shows that the first formula in (23) is

$$\alpha = (1 - c)x - \tfrac{1}{2}cx^3 + \cdots. \tag{24}$$

On differentiating this relation,

$$d\alpha/dx = (1 - c) - 3cx^2/2 + \cdots, \tag{25}$$

it is seen that, if $c \neq 1$, then $d\alpha/dx \neq 0$ for small $|x|$ and that, if $c = 1$, then $d\alpha/dx \lesseqgtr 0$ according as $|x| \gtreqless 0$. Hence, for any fixed c, α is a strictly monotone function of x (for small $|x|$). This corresponds to the fact that, in the case at hand, (12) is a one-to-one mapping of (a small) D for every c.

If γ in (23) is written as a power series in x, one obtains

$$\gamma = c + \tfrac{1}{2}(1 - c)x^2 + (1/4 - 5c/8)x^4 + \cdots. \tag{26}$$

If $c \neq 1$, then the inversion of (24) gives a power series $x = \alpha/(1 - c) + \cdots$ in α. Thus the corresponding parallel curve $\gamma = \gamma(\alpha)$ of (22) becomes

$$\gamma = c + \tfrac{1}{2}\alpha^2/(1 - c) + \cdots, \qquad (c \neq 1), \tag{27}$$

by (26). If $c = 1$, then the inversion of (24) gives

$$x = x(a) = (-2a)^{1/3} + o(|a|^{1/3}),$$

as $a \to 0$, and the corresponding parallel curve has an equation of the type

$$(28) \qquad \gamma = 1 - (3/8)(-2a)^{4/3} + o(|a|^{4/3}), \qquad (c = 1).$$

Since the parallel curve (28) clearly is not of class C^2, the resulting parallel surface of S, obtained by rotating (28) about the γ-axis, is not of class C^2. This proves (V).

As in the case of a sphere, the direction of the concavity of the parallel surface (cf. (27)) is reversed as c passes through a "critical value."

Appendix.*

Let $\mathfrak{S}$ be a (u^1, u^2)-manifold which is topologically equivalent to the sphere $x^2 + y^2 + z^2 = 1$ and carries a metric

$$(1) \qquad ds^2 = g_{\alpha\beta}(u^1, u^2)\, du^\alpha du^\beta,$$

where $\det g_{ik} > 0$, $g_{ii} > 0$, $g_{12} = g_{21}$. If the tensor g_{ik} is of class C^1 (without being of class C^2), then the existence of a continuous curvature $K = K(u^1, u^2)$ will be meant in the "integrated" sense defined by Weyl ([4], p. 43). He has stated (p. 44) the following existence theorem (along with the corresponding uniqueness theorem):

(*) Let there be given on $\mathfrak{S}$ a metric tensor (g_{ik}) which is of class C^1 and such as to possess a curvature $K = K(u^1, u^2)$ which is of class C^0 (i. e., continuous) and positive on $\mathfrak{S}$. Then there exists in the X-space, where $X = (x, y, z)$, a closed convex surface $X = X(u^1, u^2)$ of class C^2 on which the metric (1) becomes the first fundamental form, that is, $ds^2 = |dX(u, v)|^2$.

When writing his classical paper [4], Weyl was well aware that it contains certain gaps. Most of these have been filled in the meantime. But as far as we know, the latter developments have failed to lead to the existence theorem denoted by (*) above, and the following remarks would indicate the possibility that (*) is false. The situation is as follows:

For any positive integer n, let $(*_n)$ denote the statement which results if C^1 and C^0 in the assumptions, and C^2 in assertion, of (*) are replaced by the respective classes C^n, C^{n-1}, C^{n+1} (so that (*) is identical with the case $n = 1$ of $(*_n)$). We shall prove that $(*_2)$ *is false, as is* $(*_n)$ *for every* $n > 1$.

* Added March 30, 1953.

Weyl's case, $n=1$, will remain undecided. Only the case $n=2$ will be considered, since it will be clear from the proof that the method applies to $n=3,4,\cdots$ also.

In order to disprove $(*_2)$, use will be made of the following fact: As shown in [2], p. 135 (bottom), there exists on a circle $0 \leqq x^2+y^2<a^2$ a function $z=z(x,y)$ of class C^2 which is regular analytic for $0<x^2+y^2<a^2$, ceases to be of class C^3 at $(x,y)=(0,0)$, its first derivatives satisfy

$$(2) \qquad p=0 \text{ and } q=0 \text{ at } (x,y)=(0,0),$$

its third derivatives are subject to

$$(3) \qquad r_x, s_x, t_x, r_y, s_y, t_y = O(|\log(x^2+y^2)|) \text{ if } 0<x^2+y^2 \to 0,$$

finally $K(x,y)=(rt-s^2)/(1+p^2+q^2)^2$, that is, the curvature of the piece of surface $S: z=z(x,y)$, where $0 \leqq x^2+y^2<a^2$, is positive and of class C^1 (i. e., K_x, K_y exist and are continuous at $(x,y)=(0,0)$ also).

Since the metric (1) on S is

$$(4) \qquad ds^2=(1+p^2)dx^2+2pqdxdy+(1+q^2)dy^2,$$

where $(x,y)=(u^1,u^2)$ and $g_{11}=1+p^2,\cdots$, the metric tensor (g_{ik}) is of class C^1 in (x,y), simply because the function $z(x,y)$ is of class C^2. But it will be shown that, despite the fact that $z(x,y)$ is not of class C^3, the metric tensor (g_{ik}) is of class C^2 in (x,y).

First, the function $g_{11}(x,y)$, being the coefficient of dx^2 in (4), has the first partial derivatives $2pr$, $2ps$ for $0 \leqq x^2+y^2<a^2$ and the second partial derivatives

$$2r^2+2pr_x, \qquad 2rs+2pr_y, \qquad 2s^2+2ps_y$$

for $0<x^2+y^2<a^2$. Since (2) implies that $p(x,y)$ and $q(x,y)$ are $O(x^2+y^2)^{\frac{1}{2}}$ as $x^2+y^2 \to 0$, it follows from (3) and from the last formula line that the second partial derivatives of g_{11} are

$$\text{const.}+O(x^2+y^2)^{\frac{1}{2}}\log(x^2+y^2),$$

and therefore const. $+\,o(1)$. In view of well-known properties of derivatives which tend to limits, this proves that $g_{11}=g_{11}(x,y)$ has continuous second partial derivatives at $(x,y)=(0,0)$. Since the same follows if $g_{11}=1+p^2$ is replaced by $g_{12}=pq$ or $g_{22}=1+q^2$, the metric tensor (g_{ik}) is of class C^2 in (x,y) for $0 \leqq x^2+y^2<a^2$.

If the $a>0$ occurring in the definition of S is replaced by any smaller positive number, then S remains analytic on its boundary curve. It is clear from the properties of S that it is possible to choose a closed convex surface,

say U, which contains S, is of class C^∞ except at the point $(x, y) = (0, 0)$, and satisfies $K \neq 0$. Then U is an ovaloid of class C^2 on which K is positive and of class C^1 (in fact, this is true of K at the point $(x, y) = (0, 0)$ of S also). The metric tensor on U, being of class C^2 at every point of S, is of class C^2 on U. Hence, if $(*_2)$ is granted, its application to the metric tensor of U supplies the existence of an ovaloid, say V, which is of class C^3 and carries the same metric as U. Since both V and U are of class C^2 and of positive curvature, it now follows from the C^2-extension ([*1*], pp. 60-72 (A. D. Alexandrov); cf. [*5*], pp. 206-213) of Herglotz's C^3-theorem [*3*] that U and V are congruent. Consequently, U is of class C^3. But this is a contradiction, since the portion S of U is not of class C^3 at its point $(x, y) = (0, 0)$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] C. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, 1934.

[2] P. Hartman and A. Wintner, "On the third fundamental form of a surface," *American Journal of Mathematics*, vol. 75 (1953), pp. 298-334.

[3] H. Lewy, "A priori limitations for solutions of Monge-Ampère equations. II," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 365-374.

[4] A. Wintner, "On parallel surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 365-376.

Appendix.

[*1*] N. V. Efimov, "Qualitative problems of the theory of deformation of surfaces," *Uspehi Matematicheskih Nauk* (N. S), vol. 3 (1948), pp. 47-158; *Translation Number* 37, American Mathematical Society, 1951.

[*2*] P. Hartman and A. Wintner, "On the curvatures of a surface," *American Journal of Mathematics*, vol. 75 (1953), pp. 127-141.

[*3*] G. Herglotz, "Ueber die Starrheit der Eiflächen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 15 (1943), pp. 127-129.

[*4*] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[*5*] A. Wintner, "On isometric surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 198-214.

ON ASYMPTOTIC PARAMETRIZATIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

In what follows, the (sufficiently small) surface S will be of class C^n for some $n \geqq 2$, and it will be assumed that *the Gaussian curvature K is negative on S.* Thus, if $X = X(u^1, u^2)$ is a C^n-parametrization of S, and if $h_{ik}du^i du^k$, where $h_{ik} = h_{ik}(u^1, u^2)$, denotes the second fundamental form of S in this parametrization, then

$$(1) \qquad h_{ik}(u^1, u^2)\, du^i du^k = 0$$

determines a unique pair of distinct directions $du^1 : du^2$, the asymptotic directions, through every point (u^1, u^2) (since

$$(2) \qquad \det h_{ik} < 0$$

in view of $K < 0$), and the three functions $h_{ik}(u^1, u^2)$ are of class C^{n-2}, as is the function $K(u^1, u^2)$.

If the parametrization $S: X = X(u^1, u^2)$ is such as to make both functions $h_{ii}(u^1, u^2)$ vanish identically, that is, if both $u^1 =$ const. and $u^2 =$ Const. are solutions of (1) for arbitrary values of the constants, then $X = X(u^1, u^2)$ will be called an asymptotic parametrization of S. In this terminology, Corollary 1 in [2] can be stated as follows:

(i) *If S possesses an asymptotic C^n-parametrization, where $n \geqq 2$, then it possesses C^{n+1}-parametrizations.*

The following considerations deal with the converse of the question answered by this theorem, in the sense that the existence of an *asymptotic* parametrization of a specified degree of smoothness will be discussed under the assumption that a given degree of smoothness is assumed in *some* parametrization. All that is straightforward in this direction is the following fact:

(ii) *If S possesses a C^m-parametrization, where $m \geqq 3$, then it possesses asymptotic C^{m-2}-parametrizations.*

Remark. If $m = 3$ (that is, if the asymptotic parametrization which is claimed by (ii) is not of class C^2), then the coefficient functions h_{ik} of (1)

* Received October 17, 1952.

belonging to the C^{m-2}-parametrization cannot be calculated directly and must be meant as having been determined first in the C^m-parametrization and then transformed, by means of the rule of doubly covariant tensors and a C^1-transformation $(u^1, u^2) \rightarrow (v^1, v^2)$ of positive Jacobian, to the asymptotic parameters. A corresponding proviso holds for $m = 2$ in (iii) and (iv), below.

Proof of (ii). Let $X = X(u^1, u^2)$, where

$$(3) \qquad (u^1)^2 + (u^2)^2 < a^2,$$

be a C^m-parametrization of S near a point, say the point $X = (0, 0, 0)$, where $X = (x, y, z)$. It is clear from (2) that, after a fixed rotation of the coordinate system (x, y, z) about $(0, 0, 0)$, it can be supposed that $h_{11}(u^1, u^2) \neq 0$ holds at $(u^1, u^2) = (0, 0)$, and therefore on the domain (3) if a is small enough. Then, if

$$(4) \qquad f_j = (-h_{12} + (-1)^j \mid \det h_{ik} \mid^{\frac{1}{2}})/h_{11}, \qquad \text{where } j = 1, 2,$$

the quadratic equation (1) splits into the two differential equations

$$(5_1) \qquad du^1/du^2 = f_1(u^1, u^2), \qquad (5_2) \qquad du^1/du^2 = f_2(u^1, u^2),$$

where both functions (4) are of class C^{m-2} on the domain (3), since all three functions $h_{ik}(u^1, u^2)$ are. But the assumption $m \geqq 3$ means that the index of the class C^{m-2} is positive. It follows therefore from a standard theorem that, if a in (3) is small enough, the differential equations (5_1), (5_2) possess on the domain (3) first integrals, say

$$(6) \qquad v^1(u^1, u^2) = v^1 = \text{const.}, \qquad v^2(u^1, u^2) = v^2 = \text{Const.},$$

such that the functions (6) of (u^1, u^2) are of class C^{m-2} and have a Jacobian $\partial(v^1, v^2)/\partial(u^1, u^2)$ which becomes $1 \neq 0$ at $(u^1, u^2) = (0, 0)$. Consequently, the substitution (6) has a local inverse, say

$$(7) \qquad u^1 = u^1(v^1, v^2), \qquad u^2 = u^2(v^1, v^2),$$

where both functions $u^j(v^1, v^2)$ are of class C^{m-2}. Accordingly, there results a local C^{m-2}-parametrization of S in terms of (v^1, v^2) if (7) is substituted into the given parametrization $X = X(u^1, u^2)$ of S. Since (6) means that the parameter curves v_1, v_2 are asymptotic curves, the proof of (ii) is complete.

It will be noted that (ii) claims less than the converse of the assertion of (i) (when $n \geqq 3$), since the difference of the two C-indices is 1 in (i) but 2 in (ii). Hence the following theorem, in which the C^{m-2} of (ii) is improved to C^{m-1} (and, incidentally, the $m \geqq 3$ of (ii) to $m \geqq 2$), is not without interest.

(iii) *If S possesses a C^m-parametrization, where $m \geqq 2$, and if the curvature $K(<0)$ is a function of class C^{m-1} in this parametrization, then S possesses asymptotic C^{m-1}-parametrizations.*

The proof of this theorem will be omitted, since it differs from the proof of Theorems (iii), (iv) in [1] (pp. 152-153, 156) only in obvious details, the "Legendre transformation" to be used being the same as loc. cit.

Although the difference of the C-indices is the same $(=1)$ in (i) as in (iii), the latter theorem is not a straightforward converse of the former, since, under the assumptions of (i), the function K is of class C^{n-2} only; so that the last assumption of (iii) imposes on K the restriction of an additional degree of continuous differentiability. The above proof of (ii) indicates that this (or some other) additional restriction cannot be omitted, that is, that the straightforward converse of (i) is false. We shall prove this negation only for the case $m=2$ (excluded in (ii)), as follows:

(iv) *Suppose that S possesses a C^m-parametrization and an asymptotic C^{m-2}-parametrization. Then, if $m=2$, it need not possess any asymptotic C^{m-1}-parametrization.*

Note that if $m>2$, then the first assumption of (iv) implies the second by virtue of (iii).

Needless to say, the second assumption of (iv), that requiring the existence of an asymptotic C^0-parametrization $(C^{m-2}=C^0)$, is meant in the following sense: All asymptotic curves of S together form two families each of which is a continuous family of Jordan arcs which cover S in a *schlicht* manner (the arcs are, incidentally, curves of class C^1, since they are defined by (5)). Thus it is *assumed* that every point of S should issue just one asymptotic curve in an asymptotic direction. That this is actually an assumption, was proved in [1], pp. 153-156, by constructing a suitable S of class C^2.

Actually, the example given loc. cit. happens to be such that one family of asymptotic curves fails to contain a continuous family of curves which covers S in a *schlicht* manner (cf. [1], the last paragraph of Section 4, p. 155). Hence *if $m \geqq 3$ in* (ii) *is replaced by $m=2$, then no interpretation of the assertion of* (ii) *remains true.*

Proof of (iv). Let $G(x), H(y)$ be functions possessing continuous second derivatives

$$g(x)=d^2G/dx^2, \qquad h(y)=d^2H/dy^2 \tag{8}$$

on respective intervals $|x| \leqq$ const., $|y| \leqq$ Const., and let

$$g(0) = 0, \qquad h(0) = 0. \tag{9}$$

Then, as in [1], p. 153, the surface

$$S\colon z = z(x, y), \text{ where } z(x, y) = G(x) + H(y) + xy, \tag{10}$$

is of class C^2 and its Gaussian curvature $K = K(x, y)$, being -1 at $(x, y) = (0, 0)$, is negative when (x, y) is in a small rectangle containing $(0, 0)$. On such a rectangle, one of the differential equations (5_1)-(5_2) of the asymptotic curves of (10) is

$$y' = f(x, y), \text{ with } y' = dy/dx, \tag{11}$$

where, as in [1], p. 153,

$$f = \{-1 + (1 - gh)^{\frac{1}{2}}\}/h \text{ if } h \neq 0 \text{ and } f = -\tfrac{1}{2}g \text{ if } h = 0, \tag{12}$$

and, correspondingly, the other of the differential equations (5_1)-(5_2) is equivalent to

$$x' = f^*(x, y), \text{ with } x' = dx/dy, \tag{11 bis}$$

where

$$f^* = \{-1 + (1 - gh)^{\frac{1}{2}}\}/g \text{ if } g \neq 0 \text{ and } f^* = -\tfrac{1}{2}h \text{ if } g = 0. \tag{12 bis}$$

It is clear from (12), (12 bis) that both functions f, f^* of (x, y) are continuous. Note that the second case of (12) or (12 bis) occurs, by (9), at $(0, 0)$ but possibly also at points (x, y) which cluster at $(0, 0)$.

The functions $g(x)$, $h(y)$ will be chosen below so that the solutions of (11), (11 bis) are uniquely determined by their initial conditions. For small $|x|$, $|y|$, $|u|$, $|v|$, let

$$x = x(y, u), \qquad y = y(x, v) \tag{13}$$

be the solutions of (11 bis), (11), respectively, satisfying

$$x(0, u) = u, \qquad y(0, v) = v. \tag{14}$$

Both functions $x(y, u)$, $y(x, v)$ are continuous functions of two variables and the substitution (13) has a unique continuous inverse

$$u = u(x, y), \qquad v = v(x, y) \tag{15}$$

near the origin (in fact, if $y = y(x; u, v)$ is the solution of (11) satisfying $y(u; u, v) = v$, then $v(x, y) = y(0; x, y)$). Thus (15) defines a one-to-one, continuous transformation of a vicinity of $(x, y) = (0, 0)$ into a vicinity of $(u, v) = (0, 0)$. Hence, if the inverse of (15), say

$$x = x(u, v), \qquad y = y(u, v), \tag{15 bis}$$

is substituted into (10), there results for the surface S an asymptotic C^0-parametrization, say $X = X(u, v)$, in a vicinity of the point $(u, v) = (0, 0)$.

The function $y(x, v)$ occurring in (13) has, by (11), a continuous partial derivative with respect to x, namely, $f(x, y(x, v))$. The functions $g(x)$, $h(y)$ will be chosen below in such a way that $y(x, v)$ will fail to have a continuous partial derivative with respect to v at $(x, v) = (0, 0)$. Then, since the second relation in (14) implies that $y_v(0, v) = 1$, it will follow that $v = v(x, y)$ in (15) is not of class C^1 in any vicinity of $(x, y) = (0, 0)$. Consequently, $X = X(u, v)$ is not a C^1-parametrization of S (in any vicinity of $(x, y) = (u, v) = (0, 0)$). For otherwise the transformation (15), the inverse of (15 bis), were of class C^1, since, by virtue of (10), the tangent plane of S is not perpendicular to the (x, y)-plane.

If $X = X(s; t)$ is any asymptotic C^0-parametrization of S, then it results from $X = X(u, v)$ by a transformation of the form

$$(16) \qquad s = s(u), \qquad t = t(v),$$

where $s(u)$, $t(v)$ are continuous, strictly monotone functions for small $|u|$, $|v|$, respectively, and it can be supposed that $s(0) = 0$, $t(0) = 0$. No such parametrization $X(s; t) = (x(s; t), y(s; t), z(s; t))$ can be of class C^1 on any vicinity of $(s; t) = (0, 0)$. For otherwise the transformation $x = x(s; t)$, $y = y(s; t)$ has a C^1-inverse, $s = s(x, y)$, $t = t(x, y)$, with a non-vanishing Jacobian at $(x, y) = (0, 0)$. This inverse is given, in terms of (15) and the functions (16), by

$$(16\text{ bis}) \qquad s = s(u(x, y)), \qquad t = t(v(x, y)).$$

At $y = 0$ and $x = 0$, respectively, this becomes $s = s(x, 0) = s(x)$ and $t = t(0, y) = t(y)$. Hence the monotone functions (16) have non-vanishing continuous derivatives. But this is impossible, since (15) is not, but (16 bis) is, a C^1-transformation.

Thus, in order to complete the proof of (iv), it is sufficient to exhibit a pair of continuous functions $g(x)$, $h(y)$ (for small $|x|$, $|y|$, respectively) satisfying (9) and having the properties that the solutions of (11) and (11 bis) are uniquely determined by initial conditions and that $y = y(x, v)$ in (13) does not have a continuous partial derivative y_v in any vicinity of $(x, v) = (0, 0)$.

To this end, let p, q be numbers satisfying

$$(16) \qquad 0 < q < \tfrac{1}{2} \text{ and } p = q/(1 + 3q).$$

Choose

$$(17) \qquad h(y) = 0 \text{ if } y \leqq 0 \text{ and } h(y) = y^p \text{ if } y > 0.$$

Put

(18) $$a_k = 2\pi \sum_{j=k+1}^{\infty} j^{-2} \text{ for } k = 0, 1, \cdots$$

and let

(19) $$M > 1$$

be a fixed constant. In terms of q and M, define $g(x)$ as follows:

$$g(x) = 0 \text{ if } x \leqq 0,$$

(20) $$g(x) = Mk^{-q} \sin^2 k^2(t - a) \text{ if } a < t \leqq a + k^{-2}\pi,$$

$$g(x) = -M(k^{-q} - k^{-3q}) \sin^2 k^2(t - a) \text{ if } a + k^{-2}\pi < t \leqq a + 2k^{-2}\pi,$$

where $a = a_k$ and $k = 1, 2, \cdots$.

Clearly, $g(x)$ and $h(y)$ are continuous (say, for $|x| \leqq a_0$ and $|y| \leqq 1$, respectively) and satisfy (9). Since (18) implies that

(21) $$a_k \sim \text{const. } k^{-1} \text{ as } k \to \infty$$

holds for a positive constant,

(22) $$g(x) = O(x^q) \text{ as } x \to 0.$$

Furthermore, if $a = a_k$, then, as $k \to \infty$,

$$\int_0^a g\,dt = \text{const. } M \sum_{j=k+1}^{\infty} k^{-3q}k^{-2} \sim \text{Const. } Mk^{-1-3q},$$

$$\int_0^a g^2 dt = \text{const. } M^2 \sum_{j=k+1}^{\infty} k^{-2q}k^{-2} \sim \text{Const. } M^2k^{-1-2q}.$$

Also, if $(a =) a_k < x < a_{k-1}$, then

$$\int_a^x g\,dt = O(k^{-q-2}) \text{ and } \int_a^x g^2 dt = O(k^{-2q-2}).$$

It follows therefore from (21), and from the assumptions on q in (16), that there exist positive constants C_1, C_2 (independent of M) such that, as $x \to 0$,

(23) $$\int_0^x g\,dt \sim C_1 M x^{1+3q},$$ (24) $$\int_0^x g^2 dt \sim C_2 M^2 x^{1+2q}.$$

For small $|x|$, $|y|$, let (11) be written as

(25) $$y' = -\tfrac{1}{2}g + g^2h/8 + \cdots, \text{ where } y' = d/dx,$$

$g = g(x)$, $h = h(y)$ (the right-hand side of (25) is g times a power series in gh, obtained from the expansion of the $(1-gh)^{\frac{1}{2}}$ in f).

For small $|x_0|$, $|y_0|$, any initial condition

$$(26) \qquad y(x_0) = y_0$$

determines a unique solution of (11) (and/or (25)) if $x_0 < 0$, since $f(x, y) \equiv 0$ when $x < 0$; cf. the first part of (20). If $y_0 \neq 0$, then (26) determines a unique solution of (11), since $f(x, y)$ possesses a continuous partial derivative with respect to y for $y \neq 0$. If $y_0 = 0$ and $g(x_0) \neq 0$, then (11) is equivalent, for (x, y) near (x_0, y_0), to a differential equation of the type (11 bis) in which $f^*(x, y)$ possesses a continuous partial derivative with respect to x. Thus there only remains to consider the cases $y_0 = 0$, $g(x_0) = 0$.

If $x_0 = y_0 = 0$, a solution of (25) and (26) satisfies $y' = O(x^q)$, by (22). Hence $y(x) = O(x^{1+q})$ as $x \to 0$. It then follows from (20), (22) and (23) that

$$y(x) \equiv 0 \text{ if } x \leqq 0 \text{ and } y(x) \sim -\tfrac{1}{2}C_1 M x^{1+3q} \text{ as } x \to +0.$$

Thus $y(x) \leqq 0$ for small $|x|$, and so, by (17),

$$(27) \qquad y(x) = -\tfrac{1}{2}\int_0^x g\,dt \ (\leqq 0)$$

is the unique solution of (25) satisfying $y(0) = 0$.

If $g(x_0) = 0$ but $x_0 > 0$, then $|g(x)|/(x-x_0)^2$ is between two positive constants for x near x_0; cf. (20). It follows that any solution of (25) and $y(x_0) = 0$ has the property that $|y(x)|/|x-x_0|^3$ is between two positive constants for x near x_0. If $y = y_1(x)$ and $y = y_2(x)$ are two such solutions, it can be supposed that $|y_1(x)| \leqq |y_2(x)|$. Since

$$h(y_2) - h(y_1) = O(y_2 - y_1)/|y_1|^{1-p},$$

it follows that

$$(y_2 - y_1)' = O(g^2(h(y_2) - h(y_1))) = O((y_2 - y_1)|x - x_0|^{4-3(1-p)})$$

as $x \to x_0$. Since $4 - 3(1-p) > 0 > -1$ when $0 < p < 1$, standard uniqueness theorems (for example, Osgood's or Nagumo's criterion) imply that $y_2 \equiv y_1$.

Thus, when $|x_0|$, $|y_0|$ are small, (26) determines (locally) a unique solution of (11) and/or (25).

Let (11 bis) be written as

$$(28) \qquad x' = -\tfrac{1}{2}h + h^2 g/8 + \cdots, \text{ with } x' = dx/dy,$$

where $|x|$, $|y|$, are small. For small $|x_0|$, $|y_0|$, consider the initial condition

$$(29) \qquad x(y_0) = x_0.$$

This determines a unique solution of (11 bis) and/or (28) if $x_0 \neq 0$, since $f^*(x,y)$ in (11 bis) possesses a continuous partial derivative with respect to x if $x \neq 0$. If $x_0 = 0$ and $y_0 \neq 0$, then (11 bis) can be replaced, for (x,y) near (x_0, y_0), by a differential equation (11) in which $f(x,y)$ has a continuous partial derivative with respect to y. There remains therefore only the case $x_0 = y_0 = 0$. But then (28) and (29) have the unique solution $x = x(y)$ given by

$$x(y) = 0 \text{ if } y \leqq 0 \text{ and } x(y) = -\tfrac{1}{2}\int_0^y h\,dt \ (<0) \text{ if } y > 0;$$

cf. the first parts of (17) and (20). Consequently, every initial condition (29) determines (locally) a unique solution of (11 bis).

In order to complete the proof of (iv), it remains to be shown that if $y = y(x,v)$ is the solution of (11) and/or (25) satisfying $y(0,v) = v$, then $y(x,v)$ is not of class C^1 on any vicinity of $(x,v) = (0,0)$. It is clear that $y(x,v)$ is of class C^1 on every (x,v)-domain on which $y(x,v) \neq 0$. On such a domain, $y(x,v)$ has a continuous second mixed partial derivative, $y_{xv} = y_{vx}$, which can be calculated directly from (11) and/or (25).

Suppose that $y(x,v)$ is of class C^1 on a vicinity of $(x,v) = (0,0)$. Then $y_v(x,v) = 1 + o(1)$ as $(x,v) \to (0,0)$, since $y_v(0,v) \equiv 1$. Hence, when $y(t,v) \neq 0$ for $0 \leqq t \leqq x$, integration of y_{xv} with respect to x gives

$$(30) \qquad y_v(x,v) = 1 + (8^{-1} + o(1)) \int_0^x g^2(t) h_y(y(t,v))\,dt$$

for small positive x and v.

From (25),

$$(31) \qquad y(x,v) = v - \tfrac{1}{2}\int_0^x g\,dt + (8^{-1} + o(1)) \int_0^x g^2 h\,dt,$$

as $(x,v) \to (+0,+0)$. For small $v > 0$, let $x = x(v)$ be determined by the equation

$$(32) \qquad 16^{-1}(2v)^p \int_0^x g^2\,dt = v.$$

Then, according to (24), $x(v)$ satisfies $(1/16)(2v)^p C_2 M^2 x^{1+2q} \sim v$; so that there exists a constant $A > 0$ (independent of M) satisfying

$$x(v) \sim AM^{-2/(1+2q)} v^{(1-p)/(1+2q)}.$$

The definition (16) of p shows that the exponent of v is $1/(1+3q)$. Thus

$$(33) \qquad x(v) \sim AM^{-2/(1+2q)} v^{1/(1+3q)}.$$

It is clear from (23), (31) and (32) that

$$(34) \qquad y(x, v) \leqq 2v \text{ if } 0 \leqq x \leqq x(v).$$

Furthermore, for $x \leqq x(v)$, it follows from (23) that

$$\tfrac{1}{2} \int_0^x g\,dt = \tfrac{1}{2}(1 + o(1)) C_1 M x^{1+3q} \leqq \tfrac{1}{2}(1 + o(1)) C_1 A^{1+3q} M^{1-2(1+3q)/(1+2q)} v.$$

Since C_1 and A do not depend on M and since the last exponent of M is negative, it is clear that, if M is sufficiently large,

$$\tfrac{1}{2} \int_0^x g\,dt < v \text{ for } 0 \leqq x \leqq x(v).$$

Hence (31) shows that $y(t, v) > 0$ for $0 \leqq t \leqq x(v)$, and so (30) is valid for $x = x(v)$. It follows from (30) and (34) that, for $x = x(v)$,

$$y_v(x, v) \geqq 1 + (8^{-1} + o(1)) p (2v)^{p-1} \int_0^x g^2 dt$$

as $v \to +0$ and therefore, by (32),

$$y_v(x(v), v) \geqq 1 + p + o(1),$$

where $p > 0$. But this contains a contradiction, since $y_v(x, v) = 1 + o(1)$ as $(x, v) \to (0, 0)$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "On the asymptotic curves of a surface," *American Journal of Mathematics*, vol. 73 (1951), pp. 149-172.

[2] ——— and A. Wintner, "On the curvatures of a surface," *ibid.*, vol. 75 (1953), pp. 127-141.

ESSENTIAL SETS AND FIXED POINTS.*

By Barrett O'Neill.

1. **Introduction.** Let X be a topological space and let X^X be the space of (continuous) self-mappings of X furnished with the compact-open topology. M. K. Fort [1] calls a fixed point $x \in X$ of $f \in X^X$ *essential* if each map sufficiently near f has a fixed point arbitrarily near x. Fort investigated the questions of existence and recognition of fixed points in various cases, always requiring that X have the fixed point property. We continue this program with three changes: first, instead of essential *points* we consider essential *sets*; second, we search for essential sets in a specified region of X; third, we drop the requirement that X have the fixed point property.

(1.1) *A subset S of X is* essential *with respect to $f \in X^X$ if given any neighborhood V of S there is a neighborhood N of f in the compact-open topology on X^X such that each map in N has a fixed point in V.*

If an essential set reduces to a single point, then that point is fixed and is essential in the former sense. The fixed point property means that X itself is essential with respect to all its self-maps. The general problem is to determine all the essential sets of a given map $f \in X^X$. In particular we consider these two questions:

Location. Given an open set U, does U contain an essential component of the set $F(f)$ of fixed points of f?

Recognition. Given a component K of $F(f)$, is K essential?

Such questions cannot be settled by a knowledge of the essential points of f. In fact a component K of $F(f)$ isolated in a Euclidean neighborhood never contains an essential point, even though K itself is essential.

We make use of a "fixed point index" for sets similar to that of Leray [2] and others. An elementary proof is given of the existence of a suitable index when X is a finite polyhedron. Actually, using Čech theory, our methods apply to non-polyhedral spaces which are locally connected in the sense of homology.

* Received August 13, 1952.

2. The trace function. We follow the exposition of cohomology theory in [3] with a few minor changes. Let X be a finite polyhedron, and let T denote any triangulation of X obtained by subdividing the basic triangulation of X. Let $C(T)$ be the vector space of alternate cochains on T with real coefficients. For each simplex $\sigma_i \varepsilon T$ designate one of its two orientations as positive, and again denote the positively oriented simplex by σ_i. Let s_i be the (elementary) cochain such that $s_i(\sigma_j) = \delta_{ij}$. Define an inner product on $C(T)$ by the formula $(s_i, s_j) = \delta_{ij}$. Let n_i be the dimension of s_i.

If $f \varepsilon X^X$, f determines a class of cochain transformations of $C(T)$ into itself as follows: If T' is a sufficiently fine triangulation of X, there is a simplicial approximation of f inducing a cochain transformation $f' : C(T) \to C(T')$. If we follow f' by the usual subdivision transformation $\pi' : C(T') \to C(T)$, the composition function $f^+ = \pi' f'$ is called a cochain *transformation* induced by f.

If $f \varepsilon X^X$, U is an open set with no fixed points of f on its boundary, and f^+ is a cochain transformation on $C(T)$ induced by f, we define the *trace* of f^+ on U to be the integer $L(f^+, U) = \sum(-1)^{n_i}(s_i, f^+(s_i))$, where the summation extends over those elementary cochains s_i of $C(T)$ for which $\sigma_i \cap U \neq \phi$. Thus if T is sufficiently fine, we have $L(f^+, X) = L(f)$, the Lefschetz number of f.

Let F be a closed set of X and let $f \varepsilon X^X$. If $\{U_1, \cdots, U_k\}$ is a collection of open sets of X such that $\bar{U}_i \cap \bar{U}_j = \phi$ if $i \neq j$, the boundary of U_i contains no fixed points of f, and $F \subset \bigcup U_i$, then we say that $\{U_1, \cdots, U_k\}$ *partitions* F.

If T is sufficiently fine each of its simplexes contributes to the trace on at most one open set U_i. Thus we have

LEMMA 2. 1. *Let $\{U_1, \cdots, U_k\}$ partition the set $F(f)$ of fixed points of $f \varepsilon X^X$. If the triangulation T of X is sufficiently fine, and f^+ is a cochain transformation induced by f, then* $\sum_{i=1}^{k} L(f^+, U_i) = L(f)$.

It remains to show for any such U_i that $L(f^+, U_i)$ is independent of the choice of f^+ and T (provided the latter is sufficiently fine). The main step is the following elementary extension theorem:

LEMMA 2. 2. *Let $f \varepsilon X^X$ and let f_i $(i = 1, 2)$ be a simplicial approximation of f from triangulations T_i' to T_i. Let B_i be a T_i-subpolyhedron of X and A_i a T_i'-subpolyhedron such that* a) $\mathrm{Cl\,St}\, B_1 \cap \mathrm{Cl\,St}\, B_2 = \phi$ *and* b) $f_i(A_i) \subset B_i$. *Then $f_1 \mid A_1$ and $f_2 \mid A_2$ have a common extension $f' \varepsilon X^X$ which is a simplicial approximation of f from a triangulation T' to T.*

Proof. Let $C_i = \mathrm{Cl\,St}\, B_i$.[1] Since all triangulations are subdivisions of a common triangulation, we may extend $T_1 \mid C_1$ and $T_2 \mid C_2$ to a triangulation T. Similarly, extend $T_1' \mid \mathrm{Cl\,St}\, A_1$ and $T_2' \mid \mathrm{Cl\,St}\, A_2$ to a triangulation T', and let T^j denote the j-th barycentric subdivision of T' modulo $A_1 \cup A_2$ (that is, modify the usual subdivision operation by adding no new vertices to $A_1 \cup A_2$). T^j is also an extension of $T_1' \mid A_1$ and $T_2' \mid A_2$. If T^j is sufficiently fine and the vertex v is not in $\mathrm{St}\, A_i$ then $\mathrm{St}(v)$ has arbitrarily small diameter. Extend $T_1 \mid B_1$ and $T_2 \mid B_2$ to a triangulation T. Choose T^j so fine that if v is one of its vertices in X-$(\mathrm{St}\, A_1 \cup \mathrm{St}\, A_2)$, then $\mathrm{St}(v) \subset f^{-1}(St(w))$ for some vertex w of T. Let $f'v = w$. If $v \in A_i$, let $f'(v) = f_i(v)$. Finally if $v \in \mathrm{St}\, A_i - A_i$, there is a vertex v' of T_i such that $\mathrm{St}(v) \subset \mathrm{St}(v')$. Let $f'(v) = v'$. The vertex assignment above determines the desired simplicial approximation of f.

It is now easy to prove

LEMMA 2. 3. *If U_1 is an open set of X with no fixed points of f on its boundary then the trace $L(f^+, U_1)$ is independent of the choice of f^+ and T, provided T is sufficiently fine.*

Proof. Let U_2 be another open set such that $\{U_1, U_2\}$ partitions the set $F(f)$ of fixed points of f. Let f_i be a simplicial approximation of f from T_i' to T_i $(i = 1, 2)$. We must prove $L(f_1^+, U_1) = L(f_2^+, U_1)$.

Let $B_i = \mathrm{Cl\,St}\, U_i$ and let $A_i = f_i^{-1}(B_i)$. Then 2. 2 applies, so f has a simplicial approximation g from T' to T extending $f_1 \mid A_1$ and $f_2 \mid A_2$. But this implies that g^+ may be chosen so that

$$L(g^+, U_1) = L(f_1^+, U_1) \text{ and } L(g^+, U_2) = L(f_2^+, U_2).$$

If T is sufficiently fine we may combine these equations with the result of 2. 2:

$$L(f_2^+, U_1) + L(f_2^+, U_2) = L(g^+, U_1) + L(g^+, U_2)$$

to obtain $L(f_1^+, U_1) = L(f_2^+, U_1)$.

This lemma permits the following definition:

(2. 4) *If U is an open set of X with no fixed points of $f \in X^X$ on its boundary, the* trace $L(f, U)$ *of f on U is the common value of $L(f^+, U)$ for all sufficiently fine triangulations T.*

We collect the fundamental properties of the trace function in

---

[1] The operators Cl and St refer to the triangulation with which their argument is associated.

THEOREM 2. 5. *Let U and V be open sets of X whose boundaries contain no fixed points of $f \in X^X$. Then the trace function L has the properties:*

L1. *If U contains no fixed points of f, then $L(f, U) = 0$.*

L2. $L(f, U) + L(f, V) = L(f, U \cup V) + L(f, U \cap V)$.

L3. *There is a neighborhood N of f such that $g \in N$ implies (g, U) admissible and $L(f, U) = L(g, U)$.*

L4. *If $\bar{U}$ is a subpolyhedron of X and $f(\bar{U}) \subset U$, then $L(f, U) = L(f \mid \bar{U})$, the Lefschetz number of $f \mid \bar{U}$.*

L5. *If $U \cup f(U)$ is contained in a subpolyhedron of X isomorphic under a map h to a subpolyhedron of the finite polyhedron Y and if $g \in Y^Y$ is such that $gh = hf$ on U, then $L(f, U) = L(g, h(U))$.*

Proof. Properties L145 follow easily from the definition of L. It suffices to prove L2 in case $U \cap V = \phi$, for this result may then be applied to the sets $U - V$, $U \cap V$ and $V - U$. So suppose $U \cap V = \phi$ and let W be a third set such that $\{U, V, W\}$ partitions $F(f)$. Then both $\{U \cup V, W\}$ and $\{U, V, W\}$ partition $F(f)$, hence by 2. 2 we have

$$L(f, U) + L(f, V) + L(f, W) = L(f, U \cup V) + L(f, W)$$

and the result follows.

To prove L3, note that in this case maps g such that (g, U) is admissible form a neighborhood of f in X^X. It follows from 2. 1 and 2. 3 that for any sufficiently fine triangulation T there is a neighborhood N of f such that if $g \in N$ then $L(g, U) = L(g, U)$ for any choice of g. Finally, for every T there is a neighborhood P of f such that the maps in P have a common simplicial approximation and hence a common induced cochain transformation g. Choose T as above, then if $g \in N \cap P$ we have

$$L(g, A) = L(g^{+}, A) = L(f, A).$$

In the final section we prove that L is uniquely determined on the class of finite polyhedra by properties L2345.

3. Admissible sets. Let X be a space and L an integer-valued function of pairs (f, U) where U is an open set of X with no fixed points of f on its boundary. We call L a *fixed point index on* X if L has the properties L1234 of Theorem 2. 5. In this and the following section we assume that X is a space with a fixed point index L. The following properties of L are easily proved.

(3.1) *If* $L(f, U) \neq 0$, *then* U *is essential.* This is our basic criterion for the essentiality of U.

(3.2) *If* $L(f) \neq 0$ *then at least one set in any partition of* $F(f)$ *is essential.*

(3.3) *If* f_t *is a homotopy such that for each* $t \varepsilon [0,1]$ *no fixed points appear on the boundary of* U, *then* $L(f_0, U) = L(f_1, U)$.

In fact by L1, L is constant on components of the open subspace of X^X consisting of maps for which the boundary of U is free of fixed points.

(3.4) *If the symmetric difference of* U *and* V *contains no fixed points of* f, *then* $L(f, U) = L(f, V)$.

(3.5) *If* $\bar{V}$ *is a subpolyhedron of* X *such that* $\{U, V\}$ *partitions* $F(f)$ *and* $f(V) \subset V$, *then* $L(f, U) = L(f) - L(f \mid \bar{V})$.

This fact is particularly useful since it characterizes $L(f, U)$ in terms of *homological* invariants. One would like such a characterization for arbitrary pairs (f, U).

If $A \subset X$ and $f \varepsilon X^X$ let $F(f, A)$ denote the set of fixed points of f in A. We say that (f, A)—or merely A when f is clear from the context—is *admissible* if $F(f, A)$ is open and closed in the set of all fixed points of f.

An open set with no fixed points on its boundary, an isolated fixed point, any isolated component of $F(f)$ is admissible. The class of admissible sets of a given map forms a Boolean algebra under the usual operations. The fixed point index L may easily be extended to all admissible pairs of X. In fact if (f, A) is admissible and U and V are neighborhoods of A whose closures are disjoint from $F(f, X - A)$, then the symmetric difference of U and V is free of fixed points, so $L(f, U) = L(f, V)$. Thus we may define $L(f, A)$ to be the common value of $L(f, U)$ for such U. With minor changes, properties L1245 and 3.1245 continue to hold for L a function on admissible pairs. Also

(3.6) *If* $L(f, A) \neq 0$ *and each component of* $F(f, A)$ *is isolated, then at least one component of* $F(f, A)$ *is essential.*

(3.7) *If* K *is an isolated component of* $F(f)$ *and* $L(f, K) \neq 0$, *then* K *is essential.*

(3.8) *If* f *has a finite number of fixed point and* $L(f) \neq 0$, *then at least one fixed point is essential.*

Briefly, when the components of $F(f)$ are isolated, a fixed point index suffices to locate and recognize essential components of $F(f)$.

4. Non-isolated components. When the components of $F(f)$ are not isolated—and hence not admissible—we show that a fixed point index still suffices to locate and recognize essential components, provided limit operations are introduced.

Let B be a Boolean algebra. We use the notation of [4] where, in particular, $\leq$ denotes the order relation on B, 0 the zero element of B, b' the complement of an element b of B. A non-empty subset Σ of B is called a *system* provided $0 \not\in \Sigma$ and for each pair $a, b \in \Sigma$ there is an element $c \in \Sigma$ such that $c \leq a \cap b$. The set of systems in B is partially ordered by defining Σ' to be finer than Σ (written $\Sigma' \geq \Sigma$) provided that if $a \in \Sigma$ there exists an element b of Σ' such that $b \leq a$.

LEMMA 4.1. *Let B be a Boolean algebra and let S be a subset of B containing a non-zero element and such that if $a \cup b \in S$ and $a \cap b = 0$, then either a or b is in S. Then S contains a system maximal in B.*

Proof. S contains a (trivial) system and any collection of systems contains a maximal system. Thus we need only show that any system in S, maximal in S, is also maximal with respect to all systems in B.

So suppose Σ and Σ^* are systems in B and S respectively, with $\Sigma \geq \Sigma^*$ and $\Sigma \not\leq \Sigma^*$. We must show that Σ^* is not maximal in S. That $\Sigma \not\leq \Sigma^*$ means there is an element $c \in \Sigma$ such that $s \not\leq c$ for all $s \in \Sigma^*$. Let Σ_1 and Σ_2 be the collection of sets $s \cap c'$ and $s \cap c$, respectively, for all $s \in \Sigma^*$. Both Σ_1 and Σ_2 are systems finer than Σ^*. For either Σ_1 or Σ_2, say Σ_1, we have $\Sigma_1 \cap S \geq \Sigma_1$. If Σ^1 denotes $\Sigma_1 \cap S$, then Σ^1 is easily seen to be a system. But $\Sigma^1 \subset S$ and $\Sigma^1 \geq \Sigma^*$. Also $\Sigma^* \not\geq \Sigma^1$, since $s \cap c \neq 0$ for all $s \in \Sigma^*$. Similar results hold if $\Sigma_2 \cap S \geq \Sigma_2$. Thus Σ^* is not maximal in S and the proof is complete.

LEMMA 4.2. *If $f \in X^X$ let B be the Boolean algebra of admissible subsets of X. Let $\{A_i\}$ be a maximal system in B such that $\bigcup A_i$ is compact and let $D = \bigcap A_i$. Then D is one component of $F(f)$ if $F(f, A_i) \neq \phi$ for each i, otherwise $D = \phi$. D is essential if and only if each A_i is essential.*

Proof. The first assertion follows from these readily established facts: 1) $F(f, D)$ is a union of components of $F(f)$, since $\{A_i\}$ is maximal. 2) $F(f, D)$ is connected, since each A_i is admissible. 3) D contains a fixed

point if each A_i contains a fixed point, since then $\{F(f, A_i)\}$ is a system of compact sets.

It remains to show that if each A_i is essential, then D is essential. Let V be a neighborhood of D. Since $\{\bar{A}_i\}$ is a system of compact sets V contains an element of $\{A_i\}$ so that V is essential. Since V is open, maps near f have fixed points in V itself, consequently D is essential.

THEOREM 4. 3. *Let $f \varepsilon X^X$ and let A be an admissible set whose closure is compact. If $L(f, A) \neq 0$, then A contains an essential component of $F(f)$.*

Proof. Choose B as in 4. 2 and let S be the set of admissible subsets A' of A for which $L(f, A') \neq 0$. The set S contains a non empty element, so 4. 1 applies, and 4. 2 completes the proof.

COROLLARY 4. 4. *If $L(f, A) \neq 0$ and $F(f, A)$ is totally disconnected, then f has an essential fixed point in A.*

For example, let f be the self-map of the unit interval for which $f(x) = x \sin^2 1/x$ if $x \neq 0$ and $f(0) = 0$. Then there is an essential point—which must be zero.

In order to recognize whether a given component K of $F(f)$ is essential, first note that K may always be expressed as the intersection of a maximal system Σ of admissible sets. Then K is essential if Σ can be chosen so that for each $A \varepsilon \Sigma$ we have $L(f, A) \neq 0$.

5. The Euclidean case. Let X be a finite polyhedron, L the trace function. In this section we consider admissible pairs (f, A), where A has a neighborhood G homeomorphic under a map h with an open subset of Euclidean n-space R^n. This Euclidean case is simpler chiefly because the notion of displacement map is applicable. If $f \varepsilon X^X$ and A is a subset of X such that $A \cup f(A) \subset G$, then the displacement map $d: A \rightarrow R^n$ is defined by $d(x) = h(f(x)) - h(x)$. Thus the study of fixed points of f is replaced by the study of zeroes of d.

Fix following notation: P is an orientable relative n-pseudomanifold [5] with boundary B consisting of the orientable $(n-1)$-pseudomanifolds $B_1, \cdots, B_k$. If $P \subset G$ and (f, P) is admissible, then the displacement map d of f maps B into $R^n - (0)$. In fact the Euclidean distance from $d(B)$ to $0 \varepsilon R^n$ is positive, so we may restrict ourselves to triangulations of R^n such that the origin is in an (open) n-simplex, always denoted by τ, which is disjoint from $d(B)$. We want to use the notion of degree of a map, hence we must specify orientations for the spaces involved. An orientation

of a space—or space modulo a subspace—is a generator of the appropriate cohomology group. We shall continue to use simplicial cohomology groups, but now with the integers as coefficients. The groups $H^n(P,B)$, $H^{n-1}(B_i)$, $H^n(R^n, R^n - \tau)$, $H^{n-1}(R^n - \tau)$, and $H^{n-1}(R^n - (0))$ are infinite cyclic. If we consider $H^{n-1}(B_i)$ as a subgroup of $H^{n-1}(B)$, the coboundary homomorphism δ^* of the pair (P,B) provides an isomorphism of $H^{n-1}(B_i)$ onto $H^n(P,B)$. We may suppose that the homeomorphism h is such that $h^{-1}(\tau) \subset P - B$, hence h induces an isomorphism

$$h^*: H^n(R^n, R^n - \tau) \rightarrow H^n(P,B).$$

The coboundary homomorphism δ^* of $(R^n, R^n - \tau)$ and the homomorphism induced by the inclusion map of $R^n - \tau$ into $R^n - (0)$ are also isomorphisms. Thus the natural orientation of the simplex $\tau \subset R^n$ determines a generator of $H^n(R^n, R^n - \tau)$ and thereby, via the isomorphisms mentioned above, orientations of $R^n - \tau$, B_i, $R^n - (0)$, and (P,B). In particular the degree q_i of d on B_i, that is, the degree of the map $d \mid B_i: B_i \rightarrow R^n - (0)$, is defined in terms of these orientations.

THEOREM 5.1. *Let the orientable relative n-pseudomanifold P be a subpolyhedron of X, and let the boundary B of P be the union of the orientable $(n-1)$-pseudomanifolds $B_1, \cdots, B_k$. If $f \in X^X$ has no fixed points on B and if $P \cup f(P) \subset G$, an n-dimensional Euclidean open set of X, then $L(f,P) = (-1)^n \sum_1^k q_i$, where q_i is the degree of d on B_i.*

Proof. Let d^* be the homomorphism of simplicial cohomology groups induced by the map $d: B \rightarrow R^n - \tau$. If ω_i and θ' denote the orientations assigned to B_i and $R^n - \tau$, then $d^*(\theta') = \Sigma q_i \omega_i$. But $\delta^*(\omega_i) = \omega$, the selected orientation of (P,B), hence $\delta^*(d^*(\theta')) = (\Sigma q_i)\omega$. Also $\delta^*(\theta') = \theta$, the selected orientation of $(R^n, R^n - \tau)$. Thus $d^*(\delta^*(\theta')) = q\omega$, where q is the degree of the map $d: (P,B) \rightarrow (R^n, R^n - \tau)$. The following diagram is commutative:

$$\begin{array}{ccc} H^n(R^n, R^n - \tau) & \xrightarrow{d^*} & H^n(P,B) \\ \uparrow \delta^* & & \uparrow \delta^* \\ H^{n-1}(R^n - \tau) & \xrightarrow{d^*} & H^{n-1}(B) \end{array}$$

where, in each case, δ^* is the coboundary homomorphism and d^* the homomorphism induced by d. Consequently $q = \Sigma q_i$. It is known (par. 3, § 3, [6]) that the map f has a simplicial approximation f' arbitrarily near f and such that only n-simplexes contain fixed points. Because of L3 we may

assume that f has this property. Hence $L(f,P) = (-1)^n \Sigma(s_j, f^+(s_j))$, where the sum extends only over simplexes which contain fixed points. If necessary we alter the positive orientation of each such simplex σ_j to agree with the selected orientation of (P,B). This does not change $L(f,P)$. Now choose a simplicial approximation d' of the displacement map d of f such that $d'(\sigma) = \tau$, the n-simplex containing the origin of R^n, if and only if σ contains a fixed point of f. Using the fact that $hf \mid Bd\sigma_j$ and $d' \mid Bd\sigma_j$ are homotopic in $R^n - (0)$ for each simplex σ_j containing a fixed point, it is easily verified that $(s_j, f^+(s_j)) = \pm 1$ depending on whether $d' : \sigma_j \to \tau$ preserves or reverses orientation. But since the simplicial map $d' : (P,B) \to (R^n, R^n - \tau)$ induces d^*, we conclude that $\Sigma(s_j, f^+(s_j)) = q$, and the result follows.

THEOREM 5.2. *Let $f \in X^X$ and let K be a component of $F(f)$ isolated in a Euclidean neighborhood G of dimension $n > 1$. Let $r_1, \cdots, r_k$ be integers such that $\Sigma r_i = L(f,K)$ and let $x_1, \cdots, x_k$ be distinct points of K. Then there is a map $f' \in X^X$ arbitrarily near f such that $x_1, \cdots, x_k$ are the only fixed points of f' in G and $L(f', x_i) = r_i$.*

Proof. Let S be a solid sphere centered at the origin of R^n. Triangulate G so that if $P = \mathrm{Cl}\,\mathrm{St}\,K$, then $P \subset d^{-1}(S)$, where d is the displacement map of f. The polyhedron P is an orientable relative n-pseudomanifold. By further subdividing P we may assume that, if B is the boundary of P and $V_i = \mathrm{St}(x_i)$, then the sets $B, \bar{V}_1, \cdots, \bar{V}_k$ are mutually disjoint. It follows that $Q = P - \bigcup V_i$ is also a relative n-pseudomanifold. Let ω be the orientation of Q selected as in the previous theorem. If C_i is the boundary of V_i, then the boundary C of Q is $B \cup \bigcup C_i$.

Now let $e_i : \bar{V}_i \to S$ be a map such that x_i is the only zero of e_i in $\bar{V}_i$ and—measured in terms of the orientation on C_i induced by that of Q—$e_i \mid C_i$ has degree $-(-1)^n r_i$. The maps $e_i \mid C_i$ and $d \mid B$ constitute a map $d_1 : C \to S - (0)$. Let θ be the orientation of $S - (0)$ corresponding to that of $R^n - (0)$. It follows from the previous theorem and our definition of d_1 that $\delta^*(d_1^*(\theta)) = (L(f,K) - \Sigma r_i)\omega = 0$. By the exactness of the cohomology sequence of (Q,C) we conclude that $d_1^*(\theta)$ is contained in the image of $H^{n-1}(Q)$ in $H^n(C)$ under the homomorphism induced by the inclusion map of C into Q.

Since $S - (0)$ is a homotopy $(n-1)$-sphere a generalization of the Hopf Extension theorem [7] applies and d_1 has an extension $d' : Q \to S - (0)$. Using the maps e_i and d_1 we obtain an extension of $f \mid X - P$ to a map $f' \in X^X$ with $x_1, \cdots, x_k$ as its only fixed points in G. Furthermore when S is sufficiently small, f' is arbitrarily near f. Finally we note that $L(f, x_i)$ is the

degree of e_i on C_i measured in terms of the orientation of C_i induced by V_i—not Q—so $L(f, x_i) = + r_i$.

The form of the theorem when $n = 1$ is obvious.

Two consequences of the theorem are noteworthy. When $L(f, K) = 0$ we may take the set $\{x_1, \cdots, x_k\}$ to be empty, so K is inessential.

COROLLARY 5. 3. *If K is a component of $F(f)$ isolated in a Euclidean neighborhood, then K is essential if and only if $L(f, K) \neq 0$.*

Next we see that such components are *minimal* among all essential sets.

COROLLARY 5. 4. *If K is a component of $F(f)$ isolated in a Euclidean neighborhood and containing more than one point, then no proper subset of K is essential.*

6. An example. For a wide class of spaces the essentiality of a subset with respect to a map depends only on the behavior of the map in a neighborhood of the set.

LEMMA 6. 1. *Let A be a subset of a polyhedron X. If the maps $f, g \varepsilon X^X$ agree on a compact neighborhood of A, then A is essential with respect to g if and only if essential with respect to f.*

Proof. Assume that A is inessential with respect to f and that f and g agree on a compact neighborhood U of A. Let $V \subset U$ be a neighborhood of A such that maps f' arbitrarily near f have no fixed points in V, and let $F \subset V$ be a compact neighborhood of A. If N is a neighborhood of g we will show that N contains a map g' with no fixed points in F. Consider for each f' (as above) the partial map g_1 which agrees with f' on F and g on $X - V$. Given a neighborhood P of $f \mid F \varepsilon X^F$ we may choose f' so near f that $f' \mid F$ and $f \mid F$ can be joined by a homotopy h_t such that for each t we have $h_t \varepsilon P$. Thus a similar homotopy joins g_1 to $g \mid F \cup (X - V)$. Since X is an ANR it follows from Borsuk's Theorem [3] that g has an extension to $g' \varepsilon X^X$. Furthermore if P is sufficiently small, g' is arbitrarily near g (always in the compact-open topology). Since it has no fixed points in F, g' is the required map.

The preceding lemma lets us use the trace function on infinite polyhedra. As an example we investigate the essentiality of sets with respect to the following self-map of the Euclidean space R^4. Let R^3 be the subspace of points (x_1, x_2, x_3, x_4) for which $x_4 = 0$. Let S^2 be the unit sphere in R^3. Let $P(y)$ be the plane determined by $y \varepsilon S^2$ and the x_4-axis. It is easy to see

there is a self-map f_y of $P(y)$ such that: a) Near y (and $-y$) points move directly away from y (and $-y$). b) In a neighborhood of the origin, f_y is a "flow" whose streamlines are rectangular hyperbolas with one asymptote the line through y and $-y$. Near this line the flow is toward the origin. c) Only $y, -y$, and the origin are fixed points. If f_y is defined congruently on every plane $P(y)$, $y \in S^2$ we obtain a self-map f of R^4. $F(f)$ consists of the origin 0, and a 2-sphere Σ. Let A be a compact neighborhood of $F(f)$. Extend $f: A \rightarrow R^4 \subset S^4$ to a self-map F of S^4. Using 6.1 we find that any subset interior to A which is essential with respect to F in S^4 is essential with respect to f in R^4.

Now we calculate $L(F, 0)$ and $L(F, \Sigma)$. Choose a neighborhood of Σ whose boundary B is topologically the product of a circle and a 2-sphere. The displacement map d on B has the same degree as the spherical representation $r: B \rightarrow S^3$ hence $L(F, \Sigma) = 2$. On the boundary S of a small solid sphere centered at the origin, d has the same degree as the antipodal map of an equitorial 2-sphere of S, hence $L(F, 0) = -1$.

Thus both components of $F(f)$ are essential and by 5.4 also minimal. In fact the essential subsets of R^4 are exactly those which contain either the origin or the fixed sphere Σ.

7. Uniqueness of the trace function. We prove that the only topologically invariant fixed point index on the class of finite polyhedra is the trace function L. Let X be a finite polyhedron. The main fact needed is the following slight generalization of a known theorem (par. 3, § 3, [6]).

THEOREM 7.1. *If $f \in X^X$ and T is a triangulation of X there is a simplicial approximation f' of f from T' to T with the property that if $\sigma \in T'$ and $\sigma \subset f(\sigma)$, then σ is not a face of another simplex of T'.*

The proof is omitted.

A map $f \in X^X$ is said to be *normal* provided each of its fixed points x has an n-cell neighborhood K with boundary S (n depending on x) such that K contains no other points of $f^{-1}(x) \cup F(f)$, and $f(S)$ contains no interior points of K. For such a point $L(f, x) = (-1)^n q$, where q is the degree of f at x. The map f' in the preceding theorem is normal, for if x is a fixed point of f' lying in the interior of the simplex σ, then $\sigma \subset f'(\sigma)$. Consequently $F(f', Bd\sigma)$ is empty, and σ is an open set of X. Thus $L(f, U) = \Sigma(-1)^{n_i} q_i$, where the summation is over the fixed points of any sufficiently close normal approximation of f. We have proved

COROLLARY 7.2. *If U is an open set of X with no fixed points of $f \varepsilon X^X$ on its boundary, then $L(f, U)$ is a topological invariant of (f, U), that is,* L5 *holds with the map h only required to be a homeomorphism.*

We now show that L is unique.

THEOREM 7.3. *Let $\mathfrak{S}$ be a set of finite polyhedra including an n-cell and an n-sphere for each $n > 0$. If M is a topologically invariant fixed point index on each $X \varepsilon \mathfrak{S}$, then $M = L$, the trace function.*

Proof. If $f \varepsilon X^X$, $U \subset X$, $X \varepsilon \mathfrak{S}$, and the boundary of U is free of fixed points we must show $M(f, U) = L(f, U)$. We claim that it suffices to calculate $M(f, \sigma)$ in this special situation:

1) σ is an n-simplex of X with $F(f, Bd\sigma) = \phi$

2) $\bar{\sigma} \cup f(\bar{\sigma}) \subset G$, an n-dimensional Euclidean open set

3) the degree b of d on the boundary S of σ is ± 1.

By the previous discussion any map has an arbitrary close approximation f each of whose fixed points is contained in such a σ. By Theorem 5.1, $L(f, \sigma) = (-1)^n b$. If the same is true for M, it will follow from L2 and L3 that M and L are identical.

To prove that $M(f, \sigma) = (-1)^n b$ we first show that for this σ, $M(f, \sigma)$ depends only on the degree b. Then we calculate M for standard maps of degree $+1$ and -1. Let f_1 be another map for which 1) and 2) above hold and whose displacement map d_1 also has degree b on S. Since $d \mid S$ and $d_1 \mid S$ have the same degree they can be joined by a homotopy $D: S \times I \rightarrow R^n - (0)$. Let V be a neighborhood of $\bar{\sigma}$ such that $V \cup f(V) \subset G$. We now define a homotopy $H: X \times I \rightarrow X$ of f.

H is defined on a closed subset of $X \times I$ by $H(x, t) = h^{-1}(h(x) + D(x, t))$ if $x \varepsilon S$ and $H(x, t) = f(x)$ if $x \not\varepsilon V$ or $t = 0$, finally $H(x, t) = f_1(x)$ if $x \varepsilon \sigma$ and $t = 1$. We may assume that G is homeomorphic to R^n, thus there is an extension of this partial map to a homotopy $H: X \times I \rightarrow X$ from f to the map f' such that $f'(x) = H(x, 1)$. No fixed points appear on S during the homotopy, consequently L3 implies that $M(f, \sigma) = M(f', \sigma)$. But $f' \mid \bar{\sigma} = f_1 \mid \bar{\sigma}$ so we have $M(f, \sigma) = M(f_1, \sigma)$.

It remains to verify $M(f, \sigma) = (-1)^n b$ for particular maps f with $b = \pm 1$. For $b = 1$, choose h so that $h(S)$ is the unit sphere centered at the origin of R^n, then choose f so that on σ, $d = f$. Now imbed $\bar{\sigma}$ in an n-sphere S^n as its northern hemisphere. There is an obvious extension

of f to a self-map F of S^n such that F is homotopic to the identity map and F is a self-map of the southern hemisphere H of S^n. Consequently,

$$M(f,\sigma) = M(F,\sigma) = L(F) - L(F \mid H) = 1 + (-1)^n - 1 = (-1)^n.$$

If $b = -1$ and n is odd, we may suppose $f(\sigma) \subset \sigma$, so

$$M(f,\sigma) = 1 = (-1)^n b.$$

If $b = -1$ and n is even, the proof is similar to the case $b = 1$, with F a self-map of an n-cell with two fixed points outside σ.

Note that the proof does not use L1.

UNIVERSITY OF CALIFORNIA
AT LOS ANGELES.

BIBLIOGRAPHY.

[1] M. K. Fort, Jr., "Essential and non-essential fixed points," *American Journal of Mathematics*, vol. 72 (1950), pp. 315-322.

[2] J. Leray, "La theorie des points fixes et ses applications en analyse," *Proc. International Congress of Math.*, vol. II, pp. 202-208.

[3] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, 1948.

[4] G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, New York, 1946.

[5] S. Lefschetz, *Algebraic Topology*, New York, 1942.

[6] P. Alexandroff and H. Hopf, *Topologie* (Chapt. XIV), Berlin, 1935.

[7] S. Eilenberg, "Cohomology and continuous mappings," *Annals of Math.*, vol. 41 (1940), pp. 231-251.

SYMMETRIZATION RESULTS FOR SOME CONFORMAL INVARIANTS.*

By James A. Jenkins.

1. In recent years considerable attention has been paid to symmetry results and methods in the theory of functions and potential theory. A good deal of this development is due to Pólya and Szegö [7]. In particular such results have been proved for the characteristic conformal invariants of quadrangles and doubly-connected domains in various situations and for various types of symmetrization. As these invariants can be expressed in terms of the extremal lengths of certain curve families associated with the domains it is natural to expect that corresponding results hold for generalizations of extremal length. We prefer to work with quantities which we may call modules and which are essentially the reciprocals of extremal lengths. These modules are either numbers or functions associated in a conformally invariant manner with certain configurations. We will not attempt here an exhaustive enumeration of all symmetry results of this type but will give several examples clarifying the method. Also we will give an interesting application to the theory of univalent functions. We will work exclusively with circular symmetrization.

For purposes of reference we will collect here the definitions of the modules with which we will have to do, together with various explanatory remarks. In order to avoid repetition we make the following conventions throughout. In each case we deal with a domain D of finite connectivity lying in the $z(=x+iy)$ plane. On D we consider the class of functions $\rho(x,y)$ which are non-negative real valued functions of integrable square over D and subject to certain auxiliary conditions to be stated explicitly in each case. On any boundary component of D there is a natural cyclic order among the prime ends and a natural sense of description. In particular we may speak of a finite number of prime ends dividing a boundary component into sides. With D also will be associated certain classes of curves. These may be Jordan curves or open arcs on D tending to a prime end on the boundary of D at either extremity. In any case we will assume them to be locally rectifiable, i. e., if a closed subarc lies in the neighborhood associated with a local uniformizing

* Received January 12, 1953.

parameter, its image in the plane of the local uniformizing parameter shall be rectifiable. This definition is conformally invariant and can be applied even in the case of Riemann surfaces. In each case the module is given by the minimum of $\int\int_D \rho^2\, dxdy$ as ρ runs over the class of functions indicated above. The minimum is attained in all cases with which we deal here, and the corresponding metric $\rho |\, dz\,|$ will be called the extremal metric. The extremal metric is always unique, apart from trivial modifications.

I. D a quadrangle (i. e., a simply-connected domain with four prime ends distingiushed on its boundary, numbered 1, 2, 3, 4 in the natural sense). Γ the class of open arcs joining the sides 12 and 34. The auxiliary condition is that, for $\gamma \varepsilon \Gamma$, $\int_\gamma \rho \,|\, dz\,|$ exists (possibly having the value $+\infty$, this will be tacitly assumed in the later cases) with $\int_\gamma \rho \,|\, dz\,| \geqq 1$. The corresponding minimum will be denoted by M and called the module of D. It is a characteristic conformal invariant of D. If the quadrangle is mapped on a rectangle R with 1, 2, 3, 4 going into the vertices A_1, A_2, A_3, A_4 of the rectangle and A_2A_3 has length 1, then A_1A_2 has length M [2], pp. 328-329.

II. D a doubly-connected domain. C the class of Jordan curves separating its boundary continua $\mathfrak{K}_1$, $\mathfrak{K}_2$ neither of which shall reduce to a point The auxiliary condition is that, for $c \,\varepsilon\, C$, $\int_c \rho \,|\, dz\,|$ exists and $\geqq 1$. The corresponding minimum will be denoted by $\mathfrak{M}$ and called the module of D. It is a characteristic conformal invariant of D. If D is mapped on the circular ring $r_1 < |\, w\,| < r_2$, then $\mathfrak{M} = (1/2\pi)\log(r_2/r_1)$.

III. D a pentagon (i. e., a simply-connected domain with five prime ends distinguished on its boundary, numbered 1, 2, 3, 4, 5 in the natural sense.) $\Gamma_1(\Gamma_2)$ the class of open arcs joining the side 12 to 34(45). The auxiliary conditions are that, for $\gamma_i \,\varepsilon\, \Gamma_i$, $\int_{\gamma_i} \rho \,|\, dz\,|$ exists and that

$$\int_{\gamma_1} \rho \,|\, dz\,| \geqq a_1, \qquad \int_{\gamma_2} \rho \,|\, dz\,| \geqq a_2,$$

where a_1, a_2 are certain two non-negative numbers not both zero. The corresponding minimum will be denoted by $M(a_1, a_2)$. It is a function of a_1, a_2 called the module of D. Its existence is proved in [2], pp. 330-333.

IV. D a doubly-connected domain with one distinguished interior point P, neither of its boundary continua $\mathfrak{K}_1$, $\mathfrak{K}_2$ reducing to a point. Let C_1 denote the class of Jordan curves lying in D and separating $\mathfrak{K}_1$ from $\mathfrak{K}_2$ and P, let C_2 denote the class of Jordan curves lying in D and separating $\mathfrak{K}_2$ from $\mathfrak{K}_1$ and P. The auxiliary conditions are that, for $c_i \varepsilon C_i$, $(i = 1, 2)$, $\int_{c_i} \rho \mid dz \mid$ exists and that $\int_{c_1} \rho \mid dz \mid \geqq a_1$, $\int_{c_2} \rho \mid dz \mid \geqq a_2$, where a_1, a_2 are certain two non-negative numbers not both zero. The corresponding minimum will be denoted by $\mathfrak{M}(a_1, a_2)$. It is a function of a_1, a_2, called the module of D. The existence of this quantity is not given explicitly in [2] but follows from the pentagon problem in the same way that the existence of a module for a triply-connected domain follows from the hexagon problem [2], p. 348.

V. D a triply-connected domain none of whose boundary continua $\mathfrak{K}_1$, $\mathfrak{K}_2$, $\mathfrak{K}_3$ reduces to a point. Let $C_1(C_2, C_3)$ denote the class of Jordan curves lying in D and separating $\mathfrak{K}_1(\mathfrak{K}_2, \mathfrak{K}_3)$ from $\mathfrak{K}_2$ and $\mathfrak{K}_3(\mathfrak{K}_3$ and $\mathfrak{K}_1$, $\mathfrak{K}_1$ and $\mathfrak{K}_2)$. The auxiliary conditions are that, for $c_i \varepsilon C_i$ $(i = 1, 2, 3)$, $\int_{c_i} \rho \mid dz \mid$ exists and

$$\int_{c_1} \rho \mid dz \mid \geqq a_1, \qquad \int_{c_2} \rho \mid dz \mid \geqq a_2, \qquad \int_{c_3} \rho \mid dz \mid \geqq a_3,$$

where a_1, a_2, a_3 are certain three non-negative numbers not all zero. The corresponding minimum will be denoted by $\mathfrak{M}(a_1, a_2, a_3)$ and called the module of D. It is a function of a_1, a_2, a_3. Its existence for all values of a_1, a_2, a_3 is proved in [2], pp. 336-342.

2. LEMMA 1. *Let D be a doubly-connected domain in the z-plane $(z = x + iy)$ with boundary continua $\mathfrak{K}_1$ and $\mathfrak{K}_2$. Let $\omega(x, y)$ be the harmonic measure of $\mathfrak{K}_1$ relative to D. Let $D(\omega)$ denote the Dirichlet integral of ω over D. Then the module of D is equal $1/D(\omega)$.*

Since both the module and the Dirichlet integral are conformally invariant we may suppose D to be the circular ring $r_1 < \mid z \mid < r_2$ ($\mathfrak{K}_1$ being $\mid z \mid = r_2$). In this case $\omega = (\log \mid z \mid / r_1)/(\log r_2/r_1)$. Direct calculation gives $D(\omega) = 2\pi/(\log r_2/r_1)$, i. e., the reciprocal of the module.

LEMMA 2. *Let D be a quadrangle such that the sides are piecewise smooth curves, lying in the z-plane $(z = x + iy)$. Let $u(x, y)$ be a function harmonic on D, continuous on $\bar{D}$, equal to 0 on the side 12, equal to 1 on the*

side 34 *and with* $\partial u/\partial n = 0$ *on* 23 *and* 41. *Let* $D(u)$ *denote the Dirichlet integral of* u *over* D. *Then the module of* D *is equal* $1/D(u)$.

As before we can assume D to be the rectangle $0 < x < l$, $0 < y < 1$ with vertices $(0,0)$, $(l,0)$, $(l,1)$, $(0,1)$. The function in question is then $u(x,y) \equiv y$ and direct calculation gives the result.

LEMMA 3. *Let* D *be a triply-connected domain in the* $z(=x+iy)$ *plane with boundary continua* $\mathfrak{K}_1$, $\mathfrak{K}_2$, $\mathfrak{K}_3$ *as in* **V**. *Let* D_i $(i=1,2,3)$ *be doubly-connected domains lying in* D *and not overlapping such that* D_i *separates* $\mathfrak{K}_i$ *from the other two boundary continua of* D. *Let* $\mathfrak{M}_i$ *be the module of* D_i. *Then*

$$\mathfrak{M}(a_1, a_2, a_2) = \max(a_1{}^2\mathfrak{M}_1 + a_2{}^2\mathfrak{M}_2 + a_3{}^2\mathfrak{M}_3)$$

as D_1, D_2, D_3 *range over all sets of three doubly-connected domains with the above properties, it being understood that one or two of the domains* D_i *may be taken as degenerate, i. e., having module* 0.

Indeed let D_1, D_2, D_3 be such a trio of domains. Let ρ be the extremal function in the problem **V** for D and the values a_1, a_2, a_3. If any a_i vanishes, we may clearly leave it out of consideration. Otherwise in D_i, ρ/a_i is admissible (i. e., satisfies the auxiliary conditions) in the module problem for that domain. Thus $(1/a_1{}^2)\int\int_{D_i} \rho^2\, dxdy \geqq \mathfrak{M}_i$ and

$$\text{(1)} \qquad \mathfrak{M}(a_1, a_2, a_3) = \int\int_D \rho^2\, dxdy \geqq a_1{}^2\mathfrak{M}_1 + a_2{}^2\mathfrak{M}_2 + a_3{}^2\mathfrak{M}_3.$$

On the other hand, a triply-connected domain D is divided by its line of symmetry into two hexagons. From the canonical domains of [2], Fig. 3, p. 337 associated with these for various values of a_1, a_2, a_3 we can obtain representations of D by suitable identifications (see [4], p. 148). The domains so obtained are naturally divided into three doubly-connected domains for which equality is attained in (1). In the degenerate cases one or two of these doubly-connected domains will be missing.

Remark. Lemma 3 is only a very special case of an important general principle to which we will probably return in a later paper.

Let now D be a triply-connected domain for which the boundary $\mathfrak{K}_3$ is a circle $|z| = R$ and $\mathfrak{K}_1$ and $\mathfrak{K}_2$ lie interior to this circle. With D we associate a domain $\boldsymbol{D}$ obtained by circular symmetrization in the following manner. For r, $0 < r < R$, let $l_1(r)$, $l_2(r)$ be the angular Lebesgue measures of the

intersections of $\mathfrak{K}_1$ and $\mathfrak{K}_2$ with $|z| = r$. We define $\mathbf{\mathfrak{K}}_1$ as the boundary of the set: $-\frac{1}{2}l_1(r) \leqq \phi \leqq \frac{1}{2}l_1(r)$ where r, ϕ are polar coordinates and $\mathbf{\mathfrak{K}}_2$ as the boundary of the set; $\pi - \frac{1}{2}l_2(r) \leqq \phi \leqq \pi + \frac{1}{2}l_2(r)$. That these are closed sets follows by semi-continuity considerations [9], p. 598 or by approximating $\mathfrak{K}_1$, $\mathfrak{K}_2$ by analytic curves, performing the above operation and passing to the limit. Evidently $\mathbf{\mathfrak{K}}_1$ and $\mathbf{\mathfrak{K}}_2$ do not intersect one another or $|z| = R$. Then $\boldsymbol{D}$ is the domain bounded by $\mathbf{\mathfrak{K}}_1$, $\mathbf{\mathfrak{K}}_2$, $\mathfrak{K}_3$ ($= \mathbf{\mathfrak{K}}_3$, say). Let the module of $\boldsymbol{D}$ be denoted by $\mathbf{\mathfrak{M}}(a_1, a_2, a_3)$.

THEOREM 1. $\mathbf{\mathfrak{M}}(a_1, a_2, 0) \geqq \mathfrak{M}(a_1, a_2, 0)$.

First let us assume $\mathfrak{K}_1$ and $\mathfrak{K}_2$ to be analytic curves. For either one of the hexagons into which D is divided by its line of symmetry the canonical domain associated with the values $\frac{1}{2}a_1$, $\frac{1}{2}a_2$, 0 will be of type 2, 4 or 5 [2], pp. 337, 338. The corresponding canonical domain for D is obtained by taking two of these, one obtained from the other by reflection, and identifying appropriate sides. Assuming the canonical domain to be situated so that the images of $\mathfrak{K}_1$, $\mathfrak{K}_2$, $\mathfrak{K}_3$ are parallel to the imaginary axis, the real part of the function mapping D on this domain is a singled-valued function u defined on D. (The imaginary part of the mapping function will not be single-valued). The function u may be assumed to have the value 0 on $\mathfrak{K}_3$, a constant value $\beta_1 \geqq 0$ on $\mathfrak{K}_1$ and a constant value $-\beta_2 \leqq 0$ on $\mathfrak{K}_2$. The curves $\mathfrak{G}$ on which u vanishes divide D into two doubly-connected domains D_1 and D_2. D_i has $\mathfrak{K}_i$ $(i = 1, 2)$ as one boundary, the other being made up from $\mathfrak{G}$ and $\mathfrak{K}_3$. The function u is harmonic in each of D_1 and D_2 although not necessarily in D. Moreover $\mathfrak{M}(a_1, a_2, 0) = a_1^2(D(u/\beta_1; D_1))^{-1} + a_2^2(D(u/\beta_2; D_2))^{-1}$ (Lemmas 1 and 3, the second variable denotes the domain relative to which we are taking the Dirichlet integral.).

Consider the surface $u = u(x, y)$ over D completed so that $u \equiv \beta_1$ on the set bounded by $\mathfrak{K}_1$, and $u \equiv -\beta_2$ on the set bounded by $\mathfrak{K}_2$. We apply to it circular symmetrization with respect to the half-plane through the positive x-axis and perpendicular to the (x, y)-plane. In this way we obtain a function $\tilde{u}(x, y)$, still defined over $|z| \leqq R$. The set on which $\tilde{u} \equiv \beta_1$ is precisely that bounded by $\mathbf{\mathfrak{K}}_1$, the set on which $\tilde{u} \equiv -\beta_2$ is precisely that bounded by $\mathbf{\mathfrak{K}}_2$. The set on which $\tilde{u}$ vanishes divides $\boldsymbol{D}$ into two doubly-connected domains $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ corresponding to D_1 and D_2. A familiar argument [7], pp. 185, 194 shows that $D(\tilde{u}; \boldsymbol{D}_i) \leqq D(u; D_i)$, $(i = 1, 2)$. If, then, $\tilde{\omega}_i$ is the harmonic measure with respect to $\boldsymbol{D}_i$ of $\mathbf{\mathfrak{K}}_i$ we have, by Dirichlet's principle, $D(\tilde{\omega}_i; \boldsymbol{D}_i) \leqq D(\tilde{u}/\beta_i; \boldsymbol{D}_i) \leqq D(u/\beta_i; D_i)$. By Lemma 3,

$$\mathfrak{M}(a_1, a_2, 0) \geqq a_1^2(D(\tilde{\omega}_1; \boldsymbol{D}_1))^{-1} + a_2^2(D(\tilde{\omega}_2; \boldsymbol{D}_2))^{-1}$$
$$\geqq a_1^2(D(u/\beta_1; D_1))^{-1} + a_2^2(D(u/\beta_2; D_2))^{-1}.$$

Thus $\mathfrak{M}(a_1, a_2, 0) \geqq \mathfrak{M}(a_1, a_2, 0)$ as stated.

To remove the restriction that $\mathfrak{K}_1$ and $\mathfrak{K}_2$ are analytic curves we observe that any continua can be approximated from outside by such curves and the modules in question are readily seen to behave continuously under this approximation.

Let now D be a doubly-connected domain in the z-plane with the point at infinity as a distinguished interior point. Then its boundary continua $\mathfrak{K}_1$ and $\mathfrak{K}_2$ lie in a bounded part of the plane. Let $\boldsymbol{\mathfrak{K}}_1$ and $\boldsymbol{\mathfrak{K}}_2$ correspond to $\mathfrak{K}_1$ and $\mathfrak{K}_2$ under symmetrization as above. Let $\boldsymbol{D}$ be the doubly-connected domain bounded by $\boldsymbol{\mathfrak{K}}_1$ and $\boldsymbol{\mathfrak{K}}_2$, the point at infinity again being a distinguished interior point. Let $\mathfrak{M}(a_1, a_2)$ and $\boldsymbol{\mathfrak{M}}(a_1, a_2)$ be respectively the modules of D and $\boldsymbol{D}$ as in **IV**.

THEOREM 2. $\boldsymbol{\mathfrak{M}}(a_1, a_2) \geqq \mathfrak{M}(a_1, a_2)$.

This result is easily derived from Theorem 1 by regarding D and $\boldsymbol{D}$ as the limits of their intersections with a circle centered at the origin. The latter are triply-connected domains for which Theorem 1 applies and the appropriate modules tend to $\mathfrak{M}(a_1, a_2)$ and $\boldsymbol{\mathfrak{M}}(a_1, a_2)$ as the radius of the above circle tends to infinity. The result can also be obtained directly.

3. Consider now the unit circle $|z| < 1$ and the points $-r_1, 0, r_2, 1, -1$ $(0 < r_1, r_2 < 1)$ on its diameter along the real axis. Let the upper semicircle be mapped on the upper half Z-plane with these points going into Z_1, Z_2, Z_3, Z_4, Z_5 (in ascending order of magnitude on the real axis). Consider the function

$$\zeta = K \int_{Z_1}^{Z} [(Z - Z^*)^2 / \{(Z - Z_1)(Z - Z_2)^2(Z - Z_3)(Z - Z_4)(Z - Z_5)\}]^{\frac{1}{2}}\, dZ.$$

where K is a constant and Z^* is a real value in the interval $Z_4 \leqq Z^* \leqq Z_5$. For suitable choice of K this maps the upper half Z-plane on a domain bounded by rectilinear segments of the following nature. Let $A_1, A_2, A_3, A_4, A^*, A_5$ be the images of $Z_1, Z_2, Z_3, Z_4, Z^*, Z_5$. Then A_1A_2 is a half-infinite horizontal segment, A_2 being at infinity in the direction of increasing $\Re\zeta$; A_2A_3 is a half-infinite horizontal segment, the value of $\Im\zeta$ on it being larger than on A_1A_2 and $\Re\zeta$ decreasing as we go from A_2 to A_3; A_3A_4 is a vertical segment, $\Im\zeta$ decreasing as we go from A_3 to A_4; A_4A^* is a horizontal segment,

$\Re\zeta$ increasing as we go from A_4 to A^*; A^*A_5 is a horizontal segment, $\Re\zeta$ decreasing as we go from A^* to A_5; A_5A_1 is a vertical segment $\Im\zeta$ decreasing as we go from A_5 to A_1. Exceptionally A^* may coincide with A_4 or A_5 when Z^* coincides with Z_4 or Z_5 and the corresponding segment reduce to a point.

Now let the domain bounded by the segments A_1A_2, A_2A_3, A_3A_4, A_4A_5, A_5A_1 (i. e., obtained from the preceding by removing the slit penetrating to the point A^*) be mapped on the upper half w-plane with A_1, A_2, A_3, A_4, A^*, A_5 going into w_1, w_2, w_3, w_4, w^*, w_5 where $w_2 = 0$ and w_4 or $w_5 = \infty$ according as $\Re\zeta$ is smaller at A_4 or A_5. Reflecting across the segments $-1 < z < 1$ and $w_5w_1w_4$ of the real axis in the z- and w-planes we obtain a mapping of $|z| < 1$ on the w-plane minus a forked slit. We call this domain E_w. If $w_4 = \infty$, the slit runs in along the negative real axis to w_5 where it forks along curves running to w^* and $\bar{w}^*$. If $w_5 = \infty$, it runs in along the positive real axis to w_4 where it forks in a similar manner. By suitable choice of the constant K we may assume that $dw/dz = 1$ at $z = 0$.

If we extend ζ as a (non-single-valued) function of w to the whole w-plane by reflection in the various segments w_1w_2, w_2w_3, w_3w_4, w_4w_5, w_5w_1, we see at once that $d\zeta^2$ is a quadratic differential on the w-sphere, positive on $w_1w_2w_3$ and w_4w_5 and also on the slits running from $w_5(w_4)$ to w^* and $\bar{w}^*$, negative on w_5w_1 and w_3w_4. Further it has simple poles at w_1, w_3, w_4 or w_5, a simple zero at w_5 or w_4 (respectively) and a double pole at w_2. This is easily verified by introducing appropriate local uniformizing parameters at these points (as in [6]). Let us denote $d\zeta^2 = Q(w)dw^2$. The curves on which $Q(w)dw^2 > 0$ will be called trajectories. The curves on which $Q(w)dw^2 < 0$, being the orthogonal trajectories of the preceding except at critical points, will be called orthogonal trajectories.

By the way in which $Q(w)dw^2$ was obtained it is clear that the orthogonal trajectories near w_2 are closed Jordan curves. This also follows from the general theory according to which, further, they approach circular form as they shrink down to w_2 [8]. Let one of these small orthogonal trajectories be denoted by L. Let the portion of the w-plane exterior to L and slit along the real axis from w_1 to w_3 be denoted by $\mathfrak{D}$. Let $\mathfrak{C}_1$ be the class of rectifiable arcs lying in $\mathfrak{D}$ and running from L back to L separating w_3 from infinity. Let $\mathfrak{C}_2$ be the class of rectifiable arcs lying in $\mathfrak{D}$ and running from L back to L separating w_1 from infinity.

If we continue the trajectories through w^* and $\bar{w}^*$ in the direction of description from $w_5(w_4)$ to these points they run into w_2 and separate all remaining trajectories on E_w into two sets. Except for the segments w_1w_2, w_2w_3 the portions of the trajectories in one set lying in $\mathfrak{D}$ belong to $\mathfrak{C}_1$; for

the other, they belong to $\mathfrak{C}_2$. Let these respective subsets of $\mathfrak{C}_1$, $\mathfrak{C}_2$ be denoted by $\mathfrak{C}_1^*$, $\mathfrak{C}_2^*$.

In the Q-metric, $|Q(w)|^{\frac{1}{2}} |dw|$, all curves of $\mathfrak{C}_1^*$ have the same length, say a_1, and all curves of $\mathfrak{C}_2^*$ have the same length, say a_2. If $w_4 = \infty$, $a_1 \geqq a_2$ and if $w_5 = \infty$, $a_1 \leqq a_2$, equality occurring when w_4, w_5 coincide at infinity. For definiteness we will, in the remainder of this section, suppose that the first case obtains, the second being quite analogous.

We now regard the following module problem: let ρ denote a non-negative real valued function, defined and of integrable square over $\mathfrak{D}$ and such that $\int_{c_i} \rho\, |dw|\, (c_i \in \mathfrak{C}_i, \quad i = 1, 2)$ exist with $\int_{c_1} \rho\, |dw| \geqq a_1$, $\int_{c_2} \rho\, |dw| \geqq a_2$, where a_1, a_2 are the non-negative numbers given above. It is required to determine ρ so that $\int\int_{\mathfrak{D}} \rho^2\, du dv$ $(w = u + iv)$ is a minimum.

It is easily verified that $|Q(w)|^{\frac{1}{2}}$ is precisely the minimizing function. The corresponding minimum is denoted by $M^*(a_1, a_2)$ and called the module of the given configuration.

The curves of $\mathfrak{C}_1^*$ sweep out a quadrangle R_1 which has two opposite sides s_{11}, s_{13} composed of arcs of L, one further side s_{12} being the segment of the positive real axis from L to w_3 described twice, the last s_{14} being composed of the portions of the trajectories through w^* and $\bar{w}^*$ outside of L and the segment of the negative real axis from w_4 to w_5 described twice. Similarly the curves of $\mathfrak{C}_2^*$ sweep out a quadrangle R_2 which has two opposite sides s_{21} and s_{23} composed of arcs of L, one further side s_{22} being the segment of the negative real axis from w_1 to L described twice, the last s_{24} being composed of the portions of the trajectories through w^* and $\bar{w}^*$ outside of L. Let M_1 and M_2 be the modules of R_1 and R_2, in each case the class of curves Γ joining the two sides which lie on L.

As in the proof of Lemma 3 we see at once $M^*(a_1, a_2) = a_1^2 M_1 + a_2^2 M_2$.

4. Let S denote the class of functions $f(z)$, regular and univalent for $|z| < 1$ whose Taylor expansions about $z = 0$ begin

$$f(z) = z + b_2 z^2 + b_3 z^3 + \cdots + b_n z^n + \cdots.$$

It is well known that $r_1/(1 + r_1)^2 \leqq |f(-r_1)| \leqq r_1/(1 - r_1)^2$ for $0 < r_1 < 1$. In **3** we have constructed a class of functions $f_{Z^*}(z)$ contained in S and depending on the real parameter Z^*, these being the functions w carrying out the mapping indicated. As Z^* varies from Z_4 to Z_5 the function

$f_{Z^*}(z)$ varies from $z/(1+z)^2$ to $z/(1-z)^2$. It is easily seen that $f_{Z^*}(-r_1)$ varies continuously with Z^* and thus $|f_{Z^*}(-r_1)|$ takes each value in the range $(r_1/(1+r_1)^2, r_1/(1-r_1)^2)$. It is quite easy to establish that it takes each such value only once, although this is not of too great importance. It should be observed that the functions $f_{Z^*}(z)$ carry real values into real values and thus have real coefficients in their Taylor expansions about $z=0$.

Let now $f^*(z)$ be a particular function in this subclass of S and let $f(z) \varepsilon S$ be such that $|f(-r_1)| = |f^*(-r_1)|$. Then we have

THEOREM 3. $|f(r_2)| \leqq |f^*(r_2)|$ *for* $0 < r_2 < 1$.

It is readily seen to cause no loss of generality to assume $f(z)$ analytic on $|z|=1$. This function then maps $|z|<1$ on a domain, D' say, in the w'-plane bounded by an analytic curve. We can think of this as the image of E_w by a function $w'(w)$ whose Taylor expansion at $w=0$ begins as follows: $w'(w) = w + aw^2 + \cdots$.

This function maps L on a small Jordan curve L' about $w'=0$ again nearly circular and maps R_1 and R_2 onto respectively conformally equivalent quadrangles R_1' and R_2' in the w'-plane. The segment w_2w_3 goes into a curve S_1 running from 0 to $f(r_2)$ and the segment w_1w_2 goes into a curve S_2 running from $f(-r_1)$ to 0.

Suppose now we had $|f(r_2)| > |f^*(r_2)|$. Let M be a small circle, centre the origin $w'=0$, containing L' on the closed disc it bounds but touching L' from outside. If L was small enough originally, M will meet each image by $w'(w)$ of a curve of $\mathfrak{C}_1^*$ or $\mathfrak{C}_2^*$ in just two points, one near each end. Thus M will cut off from R_1' (R_2') two domains, one at each end, leaving a quadrangle R_1'' (R_2'') with a pair of opposite sides on M and the other sides lying along the corresponding sides of R_1' (R_2'). The modules M_1'', M_2'' of R_1'', R_2'' for the classes of curves joining the pair of sides lying on M in each case exceed the quantities M_1, M_2 which were the modules for R_1', R_2' by conformal invariance.

Let u_1 be a function harmonic in R_1'', continuous on its closure, with $\partial u_1/\partial n = 0$ on M, $u_1 = 1$ on S_1 and $u_1 = 0$ on the remainder of the boundary of R_1''. Let u_2 be a function harmonic in R_2'', continuous on its closure, with $\partial u_2/\partial n = 0$ on M, $u_2 = -1$ on S_2 and $u_2 = 0$ on the remainder of the boundary of R_2''. Let $u'' = u_1$ on R_1'' and its boundary, $u'' = u_2$ on R_2'' and its boundary. Let $|w'| \leqq R$ be a circle large enough to contain D' and extend u'' to vanish on the portion of this circle outside of D'.

In this way u'' is defined and continuous on the circular ring. Consider the surface $u''(\xi, \eta)$ $(w' = \xi + i\eta)$ over this ring. We apply to it circular

symmetrization with respect to the half-plane through the positive ξ-axis and perpendicular to the (ξ,η) plane. In this way we obtain a function $\tilde{u}(\xi,\eta)$ still defined over the circular ring. The set on which $1>\tilde{u}(\xi,\eta)>0$ is a quadrangle $\boldsymbol{R}_1$ with a pair of opposite sides on M and one further side along a segment Σ_1 of the positive real axis from M to a value $\geqq |f(r_2)|$, described twice. $\tilde{u}=1$ on the latter and $=0$ on the last side of $\boldsymbol{R}_1$. The set on which $0>\tilde{u}(\xi,\eta)>-1$ is a quadrangle $\boldsymbol{R}_2$ disjoint from $\boldsymbol{R}_1$ with a pair of opposite sides on M and one further side along a segment Σ_2 of the negative real axis from a value $\leqq -|f(-r_1)|$ to M. $\tilde{u}=-1$ on the latter and $=0$ on the last side of $\boldsymbol{R}_2$.

Let the modules of $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ corresponding to M_1'', M_2'' be $\boldsymbol{M}_1$, $\boldsymbol{M}_2$. Then $1/\boldsymbol{M}_i \leqq D(\tilde{u};\boldsymbol{R}_i) \leqq D(u'';R_i'')=1/M_i''$ for $i=1,2$.

The first inequality follows from Lemma 2 and the Dirichlet principle, the second follows by the usual symmetrization argument, the final equality follows by Lemma 2. Thus $\boldsymbol{M}_i \geqq M_i'' \geqq M_i$.

Consider now the domain $\mathfrak{E}$ bounded by M, Σ_1 and Σ_2 with the point at infinity regarded as a distinguished interior point. For it we regard the module problem analogous to that which in **3** defined the module $M^*(a_1,a_2)$. Here the curves of the class corresponding to $\mathfrak{C}_i$ run from M back to M and separate Σ_i from the point at infinity. The existence of the extremal metric in this case follows from the symmetry of the configuration with respect to the real axis as in [2], p. 348. The real axis divides the configuration into two symmetric halves and the corresponding problem for each is a pentagon problem. If P_1 denotes the portion of the real axis to the right of Σ_1 and P_2 denotes the portion of the real axis to the left of Σ_2, then the upper half of M, Σ_1, P_1, P_2, Σ_2 play the roles of 12, 23, 34, 45, 51. Thus Γ_1 is the class of curves joining M and P_1, Γ_2 the class of curves joining M and P_2. The appropriate auxiliary conditions are that, in the metric $\rho\,|dw'|$, curves of Γ_1 shall have length at least $\frac{1}{2}a_1$, curves of Γ_2 length at least $\frac{1}{2}a_2$. Let the module for $\mathfrak{E}$ given by this problem be denoted by $\boldsymbol{M}^*(a_1,a_2)$.

Just as in the proof of Lemma 3 we have at once that

$$\boldsymbol{M}^*(a_1,a_2) \geqq a_1{}^2\boldsymbol{M}_1 + a_2{}^2\boldsymbol{M}_2 \geqq a_1{}^2M_1 + a_2{}^2M_2 = M^*(a_1,a_2).$$

We will now show that this is in contradiction to the assumption $|f(r_2)|>|f^*(r_2)|$ provided that L be chosen small enough to begin with.

We will consider the two corresponding pentagon problems: the pentagon Π_1 having as sides the upper halves of L, s_{12}, s_{22} and the portions of the real axis to the right of s_{12} and to the left of s_{22}, the second pentagon Π_2 having as sides the upper halves of M, Σ_1, Σ_2, P_1, P_2. The function ζ maps

Π_1 on the canonical domain associated with the values $\frac{1}{2}a_1$, $\frac{1}{2}a_2$. In it the Euclidean metric is the extremal metric. Assuming Π_2 to lie also in the w-plane, ζ maps Π_2 on a subdomain of the preceding canonical domain differing from it in the following manner: whereas L goes into a vertical segment (using the normalization of ζ given in **3**), M goes into a curve joining the same horizontal lines and lying arbitrarily close to the segment when L is chosen small enough; also whereas s_{12} goes into a horizontal side, the image of Σ_1 covers this side and overlaps on the adjacent vertical side, and whereas s_{22} goes into a horizontal side, the image of Σ_2 covers this side and may overlap on the adjacent vertical side.

Now let us take in the above canonical domain a fixed vertical segment σ joining the images of s_{12} and s_{22} and lying to the left of the images of L and M. This cuts off to its left from the image of Π_1 a pentagon $\Pi_1^{(0)}$ which is in canonical form corresponding to certain numbers $a_1^{(0)}$, $a_2^{(0)}$. Let the value of the corresponding module be $M^{(0)}$. From Π_2, σ cuts off to its left a pentagon $\Pi_2^{(0)}$ whose module $\boldsymbol{M}^{(0)}$ for these numbers $a_1^{(0)}$, $a_2^{(0)}$ is strictly smaller than $M^{(0)}$. To the right of σ lies in the image of Π_1 a rectangle whose other vertical side is the image of L. Let the vertical dimension of the rectangle be l, its horizontal dimension λ. Then $\frac{1}{2}a_1 = a_1^{(0)} + \lambda$, $\frac{1}{2}a_2 = a_2^{(0)} + \lambda$. If L has been chosen small enough the image of Π_2 will contain to the right of σ a rectangle N of horizontal dimension $\lambda - \epsilon$ for any preassigned $\epsilon > 0$.

Now for the pentagon problems for Π_1 and Π_2 corresponding to the values $\frac{1}{2}a_1$, $\frac{1}{2}a_2$ and taken in their images by ζ we use the following metrics. For the image of Π_1, the Euclidean metric is the extremal metric and gives $\frac{1}{2}M^*(a_1, a_2) = M^{(0)} + l\lambda$.

For the image of Π_2 we take in $\Pi_2^{(0)}$ the extremal metric yielding the module $\boldsymbol{M}^{(0)}$ and to the right of σ we take in N the expanded Euclidean metric $\lambda(\lambda - \epsilon)^{-1} \mid d\zeta \mid$ and elsewhere the metric 0. This metric is at once seen to be admissible for the pentagon problem for Π_2 corresponding to the values $\frac{1}{2}a_1$, $\frac{1}{2}a_2$ and thus gives at least the value $\frac{1}{2}\boldsymbol{M}^*(a_1, a_2)$ but then, for ϵ small enough,

$$\tfrac{1}{2}\boldsymbol{M}^*(a_1, a_2) \leqq \boldsymbol{M}^{(0)} + (\lambda/(\lambda - \epsilon))^2(\lambda - \epsilon)l < M^{(0)} + \lambda l + 2l\epsilon < \tfrac{1}{2}M^*(a_1, a_2).$$

This contradicts the previous deduction that $\frac{1}{2}M^*(a_1, a_2) \leqq \frac{1}{2}\boldsymbol{M}^*(a_1, a_2)$ and shows that our assumption $\mid f(r_2) \mid > \mid f^*(r_2) \mid$ was false, proving the theorem.

5. THEOREM 4. *If $f(z) \varepsilon S$, and $0 < r_1 \leqq r_2 < 1$, then*

$$|f(-r_1)| + |f(r_2)| \leqq r_2/(1-r_2)^2 + r_1/(1+r_1)^2,$$

equality being attained for the function $z(1-z)^{-2}$.

Indeed, by Theorem 3, the maximum of $|f(-r_1)| + |f(r_2)|$ for r_1, r_2 fixed and f varying in the family S is attained for a function with real coefficients. For such

$$|f(-r_1)| + |f(r_2)| = r_2 + b_2 r_2^2 + \cdots + r_1 - b_2 r_1^2 + \cdots.$$

For functions with real coefficients we have $|b_n| \leqq n$ and since $r_1 \leqq r_2$

$$\begin{aligned}|f(-r_1)| + |f(r_2)| &\leqq r_2 + 2r_2^2 + 3r_2^3 + \cdots + r_1 - 2r_1^2 + 3r_1^3 + \cdots \\ &\leqq r_2/(1-r_2)^2 + r_1/(1+r_1)^2.\end{aligned}$$

The statement concerning equality is evident.

This theorem extends a result of Golusin [1], obtained by an entirely different method, namely, Löwner's parametric method.

COROLLARY 1. *If $f(z) \varepsilon S$, $0 < r_1 \leqq r_2 < 1$, then*

$$|f(-r_1 e^{i\theta})| + |f(r_2 e^{i\theta})| \leqq r_2/(1-r_2)^2 + r_1/(1+r_1)^2,$$

equality being attained for the function $z(1-ze^{-i\theta})^{-2}$.

COROLLARY 2. *If $f(z) \varepsilon S$, $0 < r < 1$ then, for $|z_1| = r$,*

$$|f(-z_1)| + |f(z_1)| \leqq 2r(1+r^2)/(1-r^2)^2,$$

equality being attained for the function $z(1-ze^{-i\theta})^{-2}$ with $\theta = \arg z_1$.

COROLLARY 3. *If $f(z) \varepsilon S$, then $|b_3| \leqq 3$.*

Indeed, if $z = re^{i\theta}$, then

$$|f(-z) + f(z)| \leqq |f(-z)| + |f(z)| \leqq 2r(1+r^2)/(1-r^2)^2.$$

Thus $r(1 + \Re(b_3 z^2) + O(r^4)) \leqq r(1 + 3r^2 + O(r^4))$ or $\Re(b_3 e^{2i\theta}) \leqq 3$.

Proper choice of θ gives $|b_3| \leqq 3$.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] G. M. Golusin, "On distortion theorems in the theory of conformal mapping," *Matematicheskii Sbornik*, vol. 60 (1946), pp. 379-389 (Russian, English summary).

[2] J. A. Jenkins, "Some problems in conformal mapping," *Transactions of the American Mathematical Society*, vol. 67 (1949), pp. 327-350.

[3] ———, "Positive quadratic differentials in triply-connected domains," *Annals of Mathematics*, vol. 53 (1951), pp. 1-3.

[4] ———, "Remarks on 'Some problems in conformal mapping,'" *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 147-151.

[5] ———, "Some results related to extremal length," to appear.

[6] ——— and D. C. Spencer, "Hyperelliptic trajectories," *Annals of Mathematics*, vol. 53 (1951), pp. 4-35.

[7] G. Pólya and G. Szegö, "Isoperimetric inequalities in mathematical physics," no. 27, *Annals of Mathematics Studies*, Princeton University Press, 1951.

[8] A. C. Schaeffer and D. C. Spencer, "Coefficient regions for schlicht functions," *American Mathematical Society Colloquium Publications*, vol. 35 (1950).

[9] V. Wolontis, "Properties of conformal invariants," *American Journal of Mathematics*, vol. 74 (1952), pp. 587-606.

THE NUMBER OF IRREDUCIBLE REPRESENTATIONS OF SIMPLE RINGS WITH NO MINIMAL IDEALS.*

By Alex Rosenberg.[1]

1. Introduction. Jacobson has called an associative ring primitive if it admits a faithful irreducible module [4]. He also showed that in case the ring possessed minimal one-sided ideals this module was unique up to an isomorphism. By examples, both Jacobson [4] and Kurosh [8] showed that the uniqueness need no longer hold if the hypothesis of a minimal one-sided ideal be dropped.

In this paper we show that a large class of simple rings with no minimal one-sided ideals always admit an infinite number of non-isomorphic irreducible modules. This is true, for example, for any simple algebraic alegbra of countable dimension with no minimal one-sided ideal and unit, and for the simple homomorph of the ring of all linear transformations on an infinite dimensional vector space.

In the last part of the paper we use our methods to complete the solution of a problem recently attacked by M. A. Naĭmark [9]. Using his results we are able to show that the ring of all completely continuous operators on a separable Hilbert space is characterized by the fact that it is a C*-algebra with a unique irreducible representation.

I should like to take this opportunity to thank Professors I. Kaplansky and D. Zelinsky for several conversations with regard to this paper. It was through the former that I became aware of Naĭmark's work.

2. Isomorphic irreducible modules. Let A be an associative ring with unit and $\mathfrak{M}$ a unitary irreducible right A-module. That is, $\mathfrak{M}$ is an additive abelian group admitting the elements of A as right operators; for all x in $\mathfrak{M}$, $x \cdot 1 = x$, 1 the unit of A; and the only submodules of $\mathfrak{M}$ are 0 and $\mathfrak{M}$. The homomorphism of A onto the ring of endomorphisms that A induces on $\mathfrak{M}$ is called an irreducible representation of A; if the homomorphism happens to be one-to-one, we speak of a faithful irreducible representation. The set

* Received February 11, 1953.

[1] Part of this work was carried out while the author was under contract to the U. S. Air Force at the University of Michigan.

of all endomorphisms of $\mathfrak{M}$ that commute with those induced by A form a division ring D. Hence $\mathfrak{M}$ may be thought of as a vector space over D and the endomorphisms induced by A as linear transformations on $\mathfrak{M}$. It is known that this ring of linear transformations is dense in $\mathfrak{M}$ over D, Theorem 6 in [4].

Two right modules $\mathfrak{M}$ and $\mathfrak{M}'$ are said to be isomorphic if there exists a one-to-one mapping σ of $\mathfrak{M}$ onto $\mathfrak{M}'$ such that $\sigma(xa) = \sigma(x)a$, x in $\mathfrak{M}$, a in A. It is well known that every irreducible right A-module is isomorphic to a quotient module A/J, where J is a maximal right ideal in A.[2] We then have

LEMMA 1. *Let A be a ring with unit. Let J and J' be two maximal right ideals in A. Then the irreducible right modules A/J and A/J' are isomorphic if and only if there is an element a in A but not in J' such that $aJ \subset J'$.*

Proof. If an isomorphism σ exists we set $\sigma(1 + J) = a + J'$. If an element a with the above properties exists, we define a mapping σ by $\sigma(x + J) = ax + J'$. It is easily seen that σ is a module homomorphism onto, with a kernel consisting of all residue classes $x + J$ such that ax lies in J'. However, the set of x with this property is a proper right ideal and so coincides with J. Thus σ is the desired isomorphism.

As an immediate consequence we have

COROLLARY 1. *Let A be a ring with unit. Then the cardinal of the set of maximal right ideals, which give rise to irreducible right modules isomorphic to a given one, is at most the cardinal of the ring.*

Proof. Let the given irreducible module be isomorphic to A/J'. Then for each J such that $A/J \cong A/J'$ there is an element a in A such that $aJ \subset J'$, a not in J'. Moreover different J's give rise to different a's, for otherwise $aA \subset J'$.

The lemma may be reformulated as follows: For convenience sake we stick to the case of a simple ring, that is one where all the irreducible representations are faithful.[3] If A is thought of as a ring of linear transformations on A/J', the maximal right ideal J' consists of all those linear transformations annihilating the vector $x = 1 + J'$. If J is a maximal right ideal such that $A/J \cong A/J'$, we see that J is the annihilator of the vector xa.

---

[2] Since A has a unit, Zorn's lemma yields the existence of maximal right ideals.

[3] A ring with unit is simple if the only two-sided ideals are 0 and the ring itself.

Conversely, let J be the annihilator of some vector $y \neq 0$ in A/J'. Since A/J' is irreducible there is an element a in A such that $xa = y$, then $aJ \subset J'$ and $A/J \cong A/J'$. Thus

LEMMA 2. *Let J be a maximal right ideal in a simple ring wtih unit A. Then all the irreducible right modules isomorphic to A/J arise from maximal right ideals which are annihilators of vectors in A/J.*

3. The ring $L - F$. Let $\mathfrak{V}$ be a vector space of dimension $\aleph \geqq \aleph_0$ over a division ring D.[4] We denote by L the ring of all linear transformations on $\mathfrak{V}$. It is known that L has a maximal two-sided ideal F, the ideal of all linear transformations whose ranges have dimension $< \aleph$; cf. [5], Theorem 2. Thus $A = L - F$ is a simple ring with unit, whose irreducible representations we now study.

The ring L may be thought of as the ring of all row finite $\aleph \times \aleph$ matrices with elements in D. Thus, as a set, L is a subset of the set of all functions from a set with $\aleph$ elements to D. Hence if the cardinal of $D = \mathfrak{d}$, $\mathfrak{l} =$ the cardinal of L, $\mathfrak{l} \leq \mathfrak{d}^{\aleph}$. In particular, if we assume that $\mathfrak{d} \leq \exp \aleph$,[5]

$$\mathfrak{l} \leq \mathfrak{d}^{\aleph} \leq (\exp \aleph)^{\aleph} = \exp \aleph^2 = \exp \aleph.$$

We shall now exhibit $\exp \exp \aleph$ distinct maximal right ideals in $A = L - F$. and so by Corollary 1 get the desired number of irreducible representations of A.

Let $B = \{x_\alpha\}$ be any basis of $\mathfrak{V}$. We consider the Boolean algebra $\mathcal{P}(B)$, of all subsets of B. To each element S of $\mathcal{P}(B)$ we associate the element e_S of L by setting

$$x_\alpha e_S = x_\alpha, \quad x_\alpha \text{ in } S, \quad x_\alpha e_S = 0, \quad x_\alpha \text{ not in } S.$$

It is easily seen that

$$e_S + e_T - e_S e_T = e_{S \cup T}, \qquad e_S e_T = e_T e_S = e_{T \cap S},$$

so that the e_S form a Boolean algebra isomorphic to $\mathcal{P}(B)$. They can also be seen to be a maximal commutative set of idempotents of L. Now by a theorem of Pospisil [10] $\mathcal{P}(B)$ contains $\exp \exp \aleph$ maximal meet ideals containing the ideal of all subsets with cardinal $< \aleph$, which we denote by $\mathcal{F}$. Let $\mathfrak{M}$ be such a maximal ideal in $\mathcal{P}(B)$. We let M be the right ideal generated in L by the e_S with S in $\mathfrak{M}$. It is easily seen that M is proper, but even more is true:

[4] That is, a Hamel basis of V has $\aleph$ elements.

[5] We write $\exp \aleph$ for the $\aleph$-th power of 2.

LEMMA 3. *The right ideal $M+F$ is proper in L.*

Proof. Suppose

$$1=\sum_1^n e_{S_i}a_i+a, \quad S_i \text{ in } M, \quad a \text{ in } F. \tag{1}$$

Then if T is the complement of $\bigcup_1^n S_i$ in B, T does not lie in $\mathfrak{M}$. By multiplying (1) on the left by e_T, we get $e_T=e_Ta$. Thus e_T is in F, T is in $\mathfrak{F}\subset\mathfrak{M}$, contradiction.

By Zorn's lemma we may then embed $M+F$ in a maximal right ideal J. Furthermore the maximal right ideals obtained in this manner from distinct ideals $\mathfrak{M}$ are also distinct. For suppose that one maximal right ideal J contained M_1 and M_2. Then for S in $\mathfrak{M}_1$, T in $\mathfrak{M}_2$ we have $B=S\cup T$, so that 1 is in $M_1+M_2\subset J$.

We have thus established the existence of $\exp\exp\aleph$ maximal right ideals in $A=L-F$. Since A has at most $\exp\aleph$ elements, it has at most $\exp\exp\aleph$ right ideals and so using Corollary 1 we get

THEOREM 1. *Let $\mathfrak{B}$ be a vector space of dimension $\aleph\geq\aleph_0$ over a division ring D. Let L be the ring of all linear transformations on $\mathfrak{B}$ and let F be the maximal two sided ideal of L, consisting of all linear transformations whose ranges have dimensions $<\aleph$. If the cardinal of $D\leq\exp\aleph$, the simple ring $L-F$ has exactly $\exp\exp\aleph$ non-isomorphic irreducible representations.*

A very similar theorem holds if L is taken to be the ring of all bounded operators on a Hilbert space. For simplicity's sake we stick to the case of a separable Hilbert space. In that case it is known that the set of all completely continuous operators forms the maximal two-sided ideal F in L, and indeed F is the only closed ideal [3], Theorem 1.4. Thus the ring $L-F$ is simple and contains a unit. Then just as before it can easily be seen that L has at most $\exp\aleph_0$ elements; and a maximal normal orthogonal set of vectors of the Hilbert space again yields $\exp\exp\aleph_0$ maximal right ideals of L containing F. Thus

THEOREM 2. *Let L be the ring of all bounded operators on a separable Hilbert space, F the set of all completely continuous operators. Then the simple ring $L-F$, considered purely algebraically, has exactly $\exp\exp\aleph_0$ non-isomorphic irreducible representations.*

Since a simple ring with a minimal one-sided ideal has a unique irreducible representation, it follows that in both cases $L-F$ is a simple ring with no minimal one-sided ideal.

4. Simple algebras of dimension $\aleph_0$. We now consider a simple algebra A with no minimal one-sided ideal over a field Φ. In order to ensure an adequate supply of idempotents we suppose that A is Zorn, i. e., *that every non-nil right ideal contains a nonzero idempotent.*[6] We shall further assume that A has a unit and that the dimension of A over $\Phi = \aleph_0$. We now proceed to construct enough maximal right ideals in A to ensure that A has $\exp \aleph_0 = \mathfrak{c}$ non-isomorphic irreducible representations. By Theorem 2. 1 of [7], A contains $\aleph_0$ orthogonal idempotents. Using Zorn's lemma, these may be embedded in a maximal commutative set of idempotents $\mathcal{B}$.

LEMMA 4. *$\mathcal{B}$ has $\aleph_0$ elements.*

Proof. Using the right regular representation we can think of A as an algebra of linear transformations on a vector space $\mathfrak{M}$ of dimension $\aleph_0$ over Φ. Now if e and f are in $\mathcal{B}$, $\mathfrak{M}ef = \mathfrak{M}fe$ so that $\mathfrak{M}ef \subset \mathfrak{M}e$. By using the Peirce decomposition we see that $\mathfrak{M}ef$ is a proper subspace unless $ef = 0$ or $e(1-f) = 0$. Since there are at most $\aleph_0$ orthogonal idempotents in A, and any strictly descending chain of subspaces of $\mathfrak{M}e$ has $\aleph_0$ elements, the lemma is proved.

By defining $e \cup f = e + f - ef$, $e \cap f = ef$ we turn $\mathcal{B}$ into a Boolean algebra with $\aleph_0$ elements and unit. $\mathcal{B}$ then has the natural ordering $e \leq f$ if and only if $ef = e$. Then

LEMMA 5. *$\mathcal{B}$ has no minimal elements.*

Proof. If e is a minimal element of $\mathcal{B}$ we consider the ring eAe. If h were an idempotent of $eAe \neq e$, it would follow that h would be in $\mathcal{B}$ too. Thus eAe has only one idempotent. But eAe is Zorn, too; [11], Lemma 2; hence eAe is a division ring and eA a minimal right ideal, contradiction.

By Stone's representation theorem it follows that $\mathcal{B}$ is isomorphic to the Boolean algebra of all open and closed subsets of a totally disconnected compact Hausdorff space $\mathfrak{S}$ with no isolated points; [13], Theorem 1. 6. Since $\mathcal{B}$ has $\aleph_0$ elements it follows that $\mathfrak{S}$ satisfies the second axiom of countability and so by the Uryson metrization theorem, it is a metric space. But this implies that $\mathfrak{S}$ is homeomorphic to the Cantor set [1], Satz VI′, p. 121. Thus $\mathfrak{S}$ has $\mathfrak{c}$ points and $\mathcal{B}$ has $\mathfrak{c}$ distinct maximal ideals. Moreover the Chinese Remainder Theorem applies to them, so that if $\mathcal{M}_1, \cdots, \mathcal{M}_{n+1}$ are maximal ideals of $\mathcal{B}$ there is an element of $\mathcal{B}$ in $\bigcap_1^n \mathcal{M}_i$ but not in $\mathcal{M}_{n+1}$.

[6] For a discussion of Zorn rings cf. [7, p. 63].

Just as in 2, we now obtain $\mathfrak{c}$ distinct maximal right ideals J of A which also have the Chinese Remainder Theorem property: $\bigcap_1^n J_i + J_{n+1} = A$. Thus if, in any irreducible representation of A, the J's are annihilators of vectors, these vectors must be linearly independent: for if x_i is annihilated by J_i and $\Sigma\alpha_i x_i = 0$, there is an element of A annihilating $x_1, \cdots, x_n$, but not x_{n+1}. However, since the dimension of A is $\aleph_0$ the dimension of any irreducible module $\mathfrak{M}$ over Φ is also $\aleph_0$, and since the commuting ring of endomorphisms D contains Φ we have that dimension of $\mathfrak{M}$ over $D = \aleph_0$. Thus it follows from Lemma 2 that at most $\aleph_0$ of the modules A/J can be isomorphic to a given one and we have

THEOREM 3. *Let A be a simple Zorn algebra of dimension $\aleph_0$. Suppose that A has a unit and no minimal one-sided ideal. Then A has at least $\mathfrak{c}$ non-isomorphic irreducible representations.*

COROLLARY 2. *Let A be a simple Zorn ring with unit, and suppose A is of dimension $\aleph_0$ over its center. Then A either satisfies the descending chain condition on right ideals or it has $\mathfrak{c}$ non-isomorphic irreducible representations.*

For if A has a minimal one-sided ideal, the presence of a unit implies the descending chain condition; [4], p. 234.

Remarks. The commuting division rings of the irreducible modules are homomorphs of subrings of A; [4], p. 236. Thus is A is an algebraic algebra over an algebraically closed field Φ, e. g., an infinite Kronecker product of 2×2 matrices with complex entries [8], these division rings are all isomorphic to Φ. Thus the representations may be non-isomorphic even if the division rings are isomorphic.

We should also like to point out that the restriction to dimension $\aleph_0$ is essential in our proof. In fact we have not succeeded in proving the existence of $\exp \aleph$ maximal right ideals in A for arbitrary $\aleph$.

5. C^*-algebras with unique irreducible representations. In a recent paper [9] Naïmark raises the following question: Let A be an irreducible C^*-algebra of bounded operators on a separable Hilbert space $\mathfrak{H}$. Suppose A admits a unique irreducible representation,[7] does this imply that A is the

[7] A representation here is a *-preserving homomorphism into the algebra of all bounded operators on a Hilbert space. Irreducibility means no invariant closed subspaces and uniqueness means that all representations are unitarily equivalent.

algebra of all completely continuous operators on $\mathfrak{H}$? Naĭmark was not able to settle the question completely; however, he did show

LEMMA 6.[8] *Every maximal commutative subalgebra of A has at most $\aleph_0$ regular maximal ideals.*

Since every closed self-adjoint commutative subalgebra of A may be embedded in a maximal one, we see by the structure theory of commutative B^*-algebras that every closed self-adjoint commutative subalgebra has at most $\aleph_0$ regular maximal ideals. Hence the space of regular maximal ideals of such algebras is a locally compact Hausdorff space with at most $\aleph_0$ points. Now

LEMMA 7. *A locally compact Hausdorff space with at most $\aleph_0$ points, contains isolated points.*[9]

Proof. By Théorème 1, § 5 of [2] such a space is of the second category. But a limit point is a set of the first category; thus isolated points exist.

But the existence of isolated points in the space of maximal ideals implies the existence of projections (self-adjoint idempotents). Hence there exists a non-empty maximal commuting set of projections B in A. Now B generates a self-adjoint closed subalgebra of A. The space of regular maximal ideals of this algebra contains at least one isolated point P. If we now take the function e which is 1 at P and 0 everywhere else, it follows that e is a minimal element of B, i. e., for every f in B, $ef = 0$ or e.

LEMMA 8. *The right ideal eA is a closed minimal ideal.*

Proof. The closure is clear. Let x be an element of eA and consider the closed subalgebra generated by xx^*. This subalgebra is in eA and by Lemmas 6 and 7 contains a projection $f = ef \neq 0$. Then $0 \neq fe$ lies in B, and since e is minimal $fe = e$. Hence $xA \supset feA = eA$, so that eA is minimal.

But then by Theorem 7. 3 of [6] it follows that A contains the ideal F of all completely continuous operators on $\mathfrak{H}$. Now if $A \neq F$, $A - F$ would still be a C^*-algebra ([6], Theorem 7. 2) and so would admit an irreducible representation [12]. This would yield an irreducible representation of A annihilating the ideal F which contradicts the uniqueness of the irreducible representations of A. Hence

[8] Naĭmark assumes that A has a unit and exactly two non-equivalent irreducible representations. However, his proofs can easily be modified to apply to the case where no unit is assumed.

[9] This lemma was pointed out to me by Professor Kaplansky.

THEOREM 4. *Let A be an irreducible C^*-algebra of bounded operators on a separable Hilbert space $\mathfrak{H}$. Suppose that A has a unique irreducible representation. Then A is the algebra of all completely continuous operators on $\mathfrak{H}$.*

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

[1] P. Alexandroff and H. Hopf, *Topologie*, vol. 1, Berlin, 1935.

[2] N. Bourbaki, "Les Structures fondamentales de l'analyse VIII (Utilisation des nombre réels en topologie générale)," *Actualités Scientifiques et Industrielles*, No. 1045, Paris, 1948.

[3] J. W. Calkin, "Two sided ideals and congruences in the ring of bounded operators on Hilbert space," *Annals of Mathematics*, vol. 42 (1941), pp. 839-873.

[4] N. Jacobson, "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.

[5] R. E. Johnson, "Equivalence rings," *Duke Mathematical Journal*, vol. 15 (1948), pp. 787-793.

[6] I. Kaplansky, "Normed algebras," *Duke Mathematical Journal*, vol. 16 (1949), pp. 399-417.

[7] ———, "Topological representations of algebras," *Transactions of the American Mathematical Society*, vol. 68 (1950), pp. 62-75.

[8] A. Kurosh, "Direct decomposition of simple rings," *Recueil Mathématique* (Matematicheskiĭ Sbornik) (N. S.), vol. 11 (1942), pp. 245-268.

[9] M. A. Naĭmark, "On a problem of the theory of rings with involution" (Russian), *Uspekhi Matematicheskikh Nauk* (N. S.), vol. 6 (1951), pp. 160-164.

[10] B. Pospisil, "On bicompact spaces," *Publications de la Faculté des Sciences de l'Université Masaryk*, No. 270, Brno, 1939.

[11] A. Rosenberg, "Subrings of simple rings with minimal ideals," *Transactions of the American Mathematical Society*, vol. 73 (1952), pp. 115-138.

[12] I. E. Segal, "Irreducible representations of operator algebras," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 73-88.

[13] M. H. Stone, "Application of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.

FRACTIONAL INTEGRATION.*

By I. I. Hirschman, Jr.[1]

1.1. Let $u(\theta) \varepsilon L^p(0, 2\pi)$ and have mean value zero, so that

$$u(\theta) \sim \Sigma' a_n e^{in\theta},$$

where $-\infty < n < \infty$ but $n \neq 0$. The fractional integral $u_\alpha(\theta)$, of order α, of $u(\theta)$, is defined by $u_\alpha(\theta) \sim \Sigma' a_n (in)^{-\alpha} e^{in\theta}$, where

$$(in)^{-\alpha} = | n |^{-\alpha} \exp(\alpha\pi i \operatorname{sgn} n/2).$$

Many interesting properties have been demonstrated in connection with fractional integration particularly by Hardy and Littlewood, see [1]-[4]. The present paper is devoted to certain results for fractional integration which are related to the work of Littlewood and Paley [5], Zygmund [8], [9] and Marcinkiewicz [6], [7]. We prove here that if $1 < p < \infty$, and $0 < \alpha < 1$, then

$$(1) \qquad A' \int_0^{2\pi} | u(\theta) |^p \, d\theta \leqq \int_0^{2\pi} d\theta \Big[\int_0^{2\pi} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) |^2 \tau^{-2\alpha-1} \, d\tau \Big]^{\frac{1}{2}p}$$
$$\leqq A'' \int_0^{2\pi} | u(\theta) |^p \, d\theta,$$

where A' and A'' are positive constants depending only on p and α. Let

$$u(\rho, \theta) \sim \sum_{-\infty}^{\infty}{}' a_n \rho^{|n|} e^{in\theta}, \qquad u_\alpha(\rho, \theta) \sim \sum_{-\infty}^{\infty}{}' a_n (in)^{-\alpha} \rho^{|n|} e^{in\theta}.$$

We also show that if $1 < p < \infty$, $-\infty < \alpha < 1$, then

$$(2) \qquad A' \int_0^{2\pi} | u(\theta) |^p \, d\theta \leqq \int_0^{2\pi} d\theta \Big[\int_0^{2\pi} (1 - \rho)^{1-2\alpha} \, | u_{\alpha-1}(\rho, \theta) |^2 \, d\rho \Big]^{\frac{1}{2}p}$$
$$\leqq A'' \int_0^{2\pi} | u(\theta) |^p \, d\theta.$$

The relations (1) and (2) are analogous $[u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau)]/\tau$ in (1) corresponding to $u_{\alpha-1}(\rho, \theta)$ in (2). Hardy and Littlewood have shown in [1], that $u \,\varepsilon\, L^p(0, 2\pi)$ implies $\| u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) \|_p = o(| \tau |)^\alpha$, $\tau \to 0$.

* Received December 16, 1952.

[1] The preparation of this paper was supported, in part, by the O. O. R. The author is a John Simon Guggenheim Memorial Fellow.

Using the above results we get some information as to the size of " o "; if $1 < p < \infty$, then

$$\int_0^{2\pi} \| u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) \|_p^r \, \tau^{-1-\alpha r} d\tau \leqq A' \| u \|_p^r, \quad r \geqq \mathrm{Max}(2, p);$$

$$\int_0^{2\pi} \| u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) \|_p^r \, \tau^{-1-\alpha r} d\tau \geqq A'' \| u \|_p^r, \quad r \leqq \mathrm{Min}(2, p).$$

1. 2. We collect here certain results which we shall need.

THEOREM 1. 2a. *Let $u(\theta) \,\varepsilon\, L^p(0, 2\pi)$, $1 < p < \infty$. We suppose $u(\theta)$ extended by periodicity to the entire real axis. Let*

$$u^\dagger(\theta) = \underset{h}{\mathrm{l.\,u.\,b.}} \, [1/h \int_\theta^{\theta+h} | u(t) | \, dt].$$

Then $\| u^\dagger(\theta) \|_p \leqq A \| u \|_p$, where A is a constant depending only on p.

This result is due to Hardy and Littlewood. A proof may be found in Zygmund [10], pp. 241-250.

Let $P(\rho, \theta)$ be the Poisson kernel

$$P(\rho, \theta) = \tfrac{1}{2\pi}(1-\rho^2)/(1 + 2\rho\cos\theta + \rho^2).$$

THEOREM 1. 2b. *Let $u(z) = \sum_1^\infty a_n z^n \,\varepsilon\, H^p(0, 2\pi)$ and*

$$\chi(\rho, \theta) = \{ \int_0^{2\pi} | u'(\rho e^{i\tau}) |^2 P(\rho, \theta-\tau) d\tau \}^{\frac{1}{2}}, \quad g^\dagger(\theta) = \{ \int_0^1 (1-\rho) \chi^2(\rho, \theta) d\rho \}^{\frac{1}{2}}.$$

Then $\| g^\dagger(\theta) \|_p \leqq A \| u \|_p$ for $1 < p < \infty$, where A is a constant depending only on p.

THEOREM 1. 2c. *Let $u(z) = \sum_1^\infty a_n z^n \,\varepsilon\, H^p(0, 2\pi)$, $\alpha > 0$ and*

$$S(\theta) = \{ \int_0^1 \int_{|\tau-\theta| \leqq \alpha(1-\rho)} | u'(\rho e^{i\tau}) |^2 \rho \, d\tau \, d\rho \}^{\frac{1}{2}}.$$

Then $\| S(\theta) \|_p \leqq A \| u \|_p$ for $1 < p < \infty$, where A is a constant depending only on α and p.

THEOREM 1. 2d. *Let $u(z) = \sum_1^\infty a_n z^n \,\varepsilon\, H^p(0, 2\pi)$, $\alpha > 0$, $\sigma > 1$ and*

$$G_\sigma(\theta) = \int_0^1 \int_{|\tau-\theta| \geqq \alpha(1-\rho)} | u'(\rho e^{i\tau}) |^2 (1-\rho)^\sigma \, | \tau - \theta |^{-\sigma} d\tau d\rho.$$

Then $\| G_\sigma(\theta) \|_p \leqq A \| u \|_p$ *for* $1 < p < \infty$, *where* A *is a constant depending only on* α, σ, *and* p.

These results are due to Littlewood and Paley, and Zygmund; see [5] and [8].

It should be pointed out that the deeper difficulties of the present subject are already overcome when these results are assumed. Many of the demonstrations here consist in making complicated though elementary reductions until a point is reached where it is possible to apply these theorems.

2.1. If $u \in L^p(0, 2\pi)$, $1 < p < \infty$, and if u has mean value zero, we define

$$\Delta(\alpha, u, \theta) = [\int_0^1 (1-\rho)^{1-2\alpha} \, | u_{\alpha-1}(\rho, \theta) |^2 \, d\rho]^{\frac{1}{2}}.$$

We shall use A for any positive constant, not necessarily the same at each appearance, which depends only on α and p.

THEOREM 2.1. *If* $1 < p < \infty$, $-\infty < \alpha < 1$, *then*

$$\| \Delta(\alpha, u, \theta) \|_p \leqq A \| u(\theta) \|_p. \tag{1}$$

It is sufficient to prove this under the assumption that $u(\rho, \theta)$ is harmonic in a region containing the unit circle in its interior. If this case has been established then, if $0 < \lambda < 1$,

$$\int_0^{2\pi} d\theta [\int_0^1 (1-\rho)^{1-2\alpha} \, | u_{\alpha-1}(\lambda\rho, \theta) |^2 \, d\rho]^{\frac{1}{2}p} \leqq A \int_0^{2\pi} | u(\lambda, \theta) |^p \, d\theta.$$

Letting $\lambda \to 1-$ and using Fatou's lemma, (1) follows.

Integrating by parts we find that

$$\Delta(\alpha, u, \theta)^2 = 2(2-2\alpha)^{-1} \int_0^1 (1-\rho)^{2-2\alpha} \, | u_{\alpha-1}(\rho, \theta) | \, d/d\rho \, | u_{\alpha-1}(\rho, \theta) | \, d\rho.$$

Since $(d/d\rho) | u_{\alpha-1}(\rho, \theta) |$ is equal in absolute value to $| (d/d\rho) u_{\alpha-1}(\rho, \theta) |$ and $d/d\rho \, u_{\alpha-1}(\rho, \theta) = u^{\sim}{}_{\alpha-2}(\rho, \theta)/\rho$, where $u^{\sim}$ is the conjugate of u, $u^{\sim}(\theta) = \sum_{-\infty}^{\infty}{}' a_n(-i \operatorname{sgn} n) e^{in\theta}$, we have, by Schwarz's inequality,

$$\Delta(\alpha, u, \theta)^2 \leqq A [\int_0^1 \rho^{-2}(1-\rho)^{1-2\alpha} \, | u_{\alpha-1}(\rho, \theta) |^2 \, d\rho]^{\frac{1}{2}} \Delta(\alpha-1, u^{\sim}, \theta).$$

Also $\int_0^1 \rho^{-2}(1-\rho)^{1-2\alpha} \, | u_{\alpha-1}(\rho, \theta) |^2 \, d\rho = I_1 + I_2$, where I_1 is the integral

from 0 to $\frac{1}{2}$ and I_2 from $\frac{1}{2}$ to 1. Using the inequality $|a_n| \leqq (2\pi)^{-1/p} \| u \|_p$, it is easy to verify that $I_1 \leqq A \| u \|_p^2$. Also $I_2 \leqq 4\Delta(\alpha, u, \theta)^2$. Thus

$$(2) \qquad \Delta(\alpha, u, \theta)^2 \leqq A[\| u \|_p + \Delta(\alpha, u, \theta)]\Delta(\alpha - 1, u^\sim, \theta).$$

It is easily deduced from this that

$$(3) \qquad \Delta(\alpha, u, \theta) \leqq A[\| u \|_p + \Delta(\alpha - 1, u^\sim, \theta)].$$

In view of this formula, if our theorem has been established for $\alpha - 1$, then it will follow for α. Indeed, suppose that it has been established for $\alpha - 1$ so that $\| \Delta(\alpha - 1, u, \theta) \|_p \leqq A \| u \|_p$. By the theorem of M. Riesz on conjugate functions $\| u^\sim \|_p \leqq A \| u \|_p$. Using these inequalities and (3), it follows that $\| \Delta(\alpha, u, \theta) \|_p \leqq A \| u \|_p$. It is thus sufficient to prove our theorem under the assumption $\alpha < -\frac{1}{2}$.

We have

$$u_{\alpha-1}(\rho, \theta) = \int_0^{2\pi} u_{-1}(\rho^{\frac{1}{2}}, t) k(\rho^{\frac{1}{2}}, \theta - t)\, dt, \qquad k(\rho, \theta) = \frac{1}{2\pi} \sum_{-\infty}^{\infty}{}' (in)^{-\alpha} \rho^{|n|} e^{in\theta}.$$

It is easily verified that $| k(\rho, \theta) | \leqq A\, | 1 - \rho e^{i\theta} |^{-1+\alpha}$ for $|\rho| < 1$. Thus

$$u_{\alpha-1}(\rho, \theta) \leqq A \int_0^{2\pi} | u_{-1}(\rho^{\frac{1}{2}}, t) |\, | 1 - \rho^{\frac{1}{2}} e^{i(\theta - t)} |^{-1+\alpha}\, dt.$$

Using Schwarz's inequality we find that

$$| u_{\alpha-1}(\rho, \theta) |^2 \leqq A \int_0^{2\pi} | u_{-1}(\rho^{\frac{1}{2}}, t) |^2\, | 1 - \rho^{\frac{1}{2}} e^{i(\theta - t)} |^{-2}\, dt \int_0^{2\pi} | 1 - \rho^{\frac{1}{2}} e^{i(\theta - t)} |^{2\alpha}\, dt.$$

Provided $\alpha < -\frac{1}{2}$, we have $\int_0^{2\pi} | 1 - \rho^{\frac{1}{2}} e^{i(\theta - t)} |^{2\alpha}\, dt \leqq A(1 - \rho^{\frac{1}{2}})^{2\alpha+1}$. Since, $P(\rho, t)$ being the Poisson kernel, $(1 - \rho^{\frac{1}{2}}) | 1 - \rho^{\frac{1}{2}} e^{i(\theta - t)} |^{-2} \leqq AP(\rho^{\frac{1}{2}}, \theta - t)$, it follows that

$$| u_{\alpha-1}(\rho, \theta) |^2 \leqq A(1 - \rho^{\frac{1}{2}})^{2\alpha} \int_0^{2\pi} | u_{-1}(\rho^{\frac{1}{2}}, t) |^2 P(\rho^{\frac{1}{2}}, \theta - t)\, dt;$$

$$\Delta(\alpha, u, \theta)^2 \leqq A \int_0^1 (1 - \rho)^{1-2\alpha} (1 - \rho^{\frac{1}{2}})^{2\alpha} d\rho \int_0^{2\pi} | u_{-1}(\rho^{\frac{1}{2}}, t) |^2 P(\rho^{\frac{1}{2}}, \theta - t)\, dt$$

$$\leqq \int_0^1 (1 - r) \int_0^{2\pi} | u_{-1}(r, t) |^2 P(r, \theta - t)\, dt\, dr.$$

That $\| \Delta(\alpha, u, \theta) \|_p \leqq A \| u \|_p$ is now a consequence of Theorem 1.2b, and M. Riesz's theorem on conjugate functions.

2.2. In this section we will establish the converse inequality.

THEOREM 2.2. *If* $1 < p < \infty$, $-\infty < \alpha < -1$, *then*

$$\| \Delta(\alpha, u, \theta) \|_p \geqq A \| u(\theta) \|_p. \tag{1}$$

Again, it is sufficient to prove this under the assumption that $u(\rho, \theta)$ is harmonic in a region containing the unit circle in its interior. For if it has been proven in this case and if $0 < \lambda < 1$, then

$$\| u(\lambda, \theta) \|_p \leqq A \| \Delta(\alpha, u(\lambda\rho, \theta), \theta) \|_p.$$

Now

$$\Delta(\alpha, u(\lambda\rho, \theta), \theta)^2 = \int_0^1 (1-\rho)^{1-2\alpha} | u_{\alpha-1}(\lambda\rho, \theta)|^2 d\rho$$

$$= \lambda^{-1} \int_0^\lambda (1-\rho\lambda^{-1})^{1-2\alpha} | u_{\alpha-1}(\rho, \theta)|^2 d\rho \leqq \lambda^{-1}\Delta(\alpha, u, \theta)^2.$$

Thus we have $\| u(\lambda, \theta)\|_p \leqq A\lambda^{-\frac{1}{2}} \| \Delta(\alpha, u, \theta)\|_p$. Letting $\lambda \to 1-$ we obtain (1).

Let $v(\rho, \theta)$ be harmonic in a region containing the unit circle in its interior and let $v(0, \theta) = 0$. We set $v(\theta) = v(1, \theta)$. Inserting the Fourier expansions of $u(\rho, \theta) v(\rho, \theta)$, etc., it is very easy to verify that

$$\int_0^{2\pi} u(\theta)v(\theta)d\theta$$

$$= \int_0^{2\pi} \int_0^1 (1-\rho)\rho^{-2}[4u_{-1}(\rho, \theta)v_{-1}(\rho, \theta) + 2u_{-1}(\rho, \theta)v^{\sim}(\rho, \theta)] d\rho d\theta.$$

Thus $\int_0^{2\pi} u(\theta) v(\theta) d\theta = \int_0^{2\pi} u_\alpha(\theta) v_{-\alpha}(\theta) d\theta$

$$= \int_0^{2\pi} \int_0^1 (1-\rho)\rho^{-2}[4u_{\alpha-1}(\rho, \theta)v_{-\alpha-1}(\rho, \theta) + 2u_{\alpha-1}(\rho, \theta)v^{\sim}_{-\alpha}(\rho, \theta)] d\rho d\theta,$$

$= I_1 + I_2$, where I_1 corresponds to the range $0 \leqq \theta \leqq 2\pi$, $0 \leqq \rho \leqq \epsilon$ and I_2 to the range $0 \leqq \theta \leqq 2\pi$, $\epsilon \leqq \rho \leqq 1$. Let us now suppose $\| v \|_q \leqq 1$. Given $a > 0$ it is possible to choose $\epsilon > 0$, depending only upon p, α and a, so that $| I_1 | \leqq a \| u \|_p$. There will then exist a constant $A(a)$ such that

$$| I_2 | \leqq A(a) \int_0^{2\pi} \int_0^1 (1-\rho) [4u_{\alpha-1}(\rho, \theta) v_{-\alpha-1}(\rho, \theta)$$

$$+ 2u_{\alpha-1}(\rho, \theta) v^{\sim}_{-\alpha}(\rho, \theta)] d\rho d\theta.$$

By Schwarz's inequality,

$$\left| \int_0^1 (1-\rho) u_{\alpha-1}(\rho, \theta) v_{-\alpha-1}(\rho, \theta) d\rho \right| \leqq \Delta(\alpha, u, \theta)\Delta(-\alpha, v, \theta).$$

By Theorem 2.1 and Hölder's inequality,

$$\left| \int_0^{2\pi} \int_0^1 (1-\rho) u_{\alpha-1}(\rho, \theta) v_{-\alpha-1}(\rho, \theta) d\rho d\theta \right| \leqq A \| \Delta(\alpha, u, \theta) \|_p .$$

Let $w = v^{\sim}{}_1$; evidently $\| w \|_q \leqq A$. Again

$$\left| \int_0^1 (1-\rho) u_{\alpha-1}(\rho, \theta) v^{\sim}{}_{-\alpha}(\rho, \theta) d\rho \right| \leqq \Delta(\alpha, u, \theta) \Delta(-\alpha, w, \theta),$$

$$\left| \int_0^{2\pi} \int_0^1 (1-\rho) u_{\alpha-1}(\rho, \theta) v^{\sim}{}_{-\alpha}(\rho, \theta) d\rho d\theta \right| \leqq A \| \Delta(\alpha, u, \theta \|_p .$$

It follows that $\left| \int_0^{2\pi} u(\theta) v(\theta) d\theta \right| \leqq a \| u \|_p + A(a) \| \Delta(\alpha, u, \theta) \|_p$. We have l. u. b. $\left| \int_0^{2\pi} u(\theta) v(\theta) d\theta \right| \geqq \frac{1}{2} \| u \|_p$, when v varies over the class of functions described above and thus $\frac{1}{2} \| u \|_p \leqq a \| u \|_p + A(a) \| \Delta(\alpha, u, \theta) \|_p$. Since a is arbitrary, this implies our theorem for $\alpha > -1$, the restriction arising from the fact that in our proof we applied Theorem 2.1 with $-\alpha$. The extension to other values of α is easily effected, using (2) of § 2.1.

For $\alpha = 0$ Theorems 2.1 and 2.2 have been established by Littlewood and Paley [5].

3.1. If $u \in L^p(0, 2\pi)$ and if u has mean value zero, we define

$$\delta(\alpha, u, \theta) = \left[\int_0^{2\pi} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) |^2 \tau^{-2\alpha-1} d\tau \right]^{\frac{1}{2}}.$$

THEOREM 3.1. *If* $1 < p < \infty$, $0 < \alpha < 1$, *then*

$$\| \delta(\alpha, u, \theta) \|_p \leqq A \| u \|_p . \tag{1}$$

Let $v(\theta) = \sum_1^\infty a_n e^{in\theta}$, $w(\theta) = \sum_{-\infty}^{-1} a_n e^{in\theta}$. We have by Riesz's theorem, see [10], pp. 147-151, $\| v \|_p \leqq A \| u \|_p$, $\| w \|_p \leqq A \| u \|_p$. Further,

$$\begin{aligned} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) |^2 \leqq 2 \, & | v_\alpha(\theta + \tau) - v_\alpha(\theta - \tau) |^2 \\ & + 2 \, | w_\alpha(\theta + \tau) - w_\alpha(\theta - \tau) |^2, \end{aligned}$$

and from this it is easily deduced that $\delta(\alpha, u, \theta) \leqq A\delta(\alpha, v, \theta) + A\delta(\alpha, w, \theta)$. Thus if our theorem were demonstrated for v and w, so that

$$\| \delta(\alpha, v, \theta) \|_p \leqq A \| v \|_p, \qquad \| \delta(\alpha, w, \theta) \|_p \leqq A \| w \|_p,$$

then (1) would follow. Thus we may assume that $u(\rho, \theta) = \sum_1^\infty a_n \rho^n e^{in\theta}$.

We write $u(z)$, $z = \rho e^{i\theta}$ for $u(\rho, \theta)$. Further it is no restriction to suppose that $u(z)$ is analytic in a region containing the unit circle in its interior.

Let $\rho_T = 1 - \tau/(4\pi)$. We have

$$\begin{aligned} u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) = {} & \tfrac{1}{2} \int_{\rho_T e^{i(\theta+T)}}^{e^{i(\theta+T)}} [z - e^{i(\theta+\tau)}]^2 u'''_\alpha(z)\,dz \\ & + \tfrac{1}{2} \int_{\rho_T e^{i\theta}}^{\rho_T e^{i(\theta+T)}} [z - e^{i(\theta+\tau)}]^2 u'''_\alpha(z)\,dz \\ & + \tfrac{1}{2} u''_\alpha(\rho_T e^{i\theta}) \{-[\rho_T e^{i\theta} - e^{i(\theta+\tau)}]^2 + [\rho_T e^{i\theta} - e^{i(\theta-\tau)}]^2\} \\ & + u'_\alpha(\rho_T e^{i\theta}) \{[\rho_T e^{i\theta} - e^{i(\theta+\tau)}] - [\rho_T e^{i\theta} - e^{i(\theta-\tau)}]\} \\ & - \tfrac{1}{2} \int_{\rho_T e^{i(\theta-T)}}^{e^{i(\theta-T)}} [z - e^{i(\theta-\tau)}]^2\, u'''_\alpha(z)\,dz \\ & - \tfrac{1}{2} \int_{\rho_T e^{i(\theta-T)}}^{\rho_T e^{i\theta}} [z - e^{i(\theta-\tau)}]^2\, u'''_\alpha(z)\,dz, \\ u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) = {} & I_1 + I_2 + I_3 + I_4 + I_5 + I_6. \end{aligned}$$

Consider I_1. We have

$$\begin{aligned} |I_1| &\leqq \int_{\rho_T}^{1} |u'''_\alpha(\rho e^{i(\theta+\tau)})|\,(1-\rho)^2 d\rho, \\ |I_1|^2 &\leqq \int_{\rho_T}^{1} |u'''_\alpha(\rho e^{i(\theta+\tau)})|^2 (1-\rho)^{5-\alpha} d\rho \int_{\rho_T}^{1} (1-\rho)^{-1+\alpha} d\rho, \\ &\leqq A\tau^\alpha \int_{\rho_T}^{1} |u'''_\alpha(\rho e^{i(\theta+\tau)})|^2 (1-\rho)^{5-\alpha} d\rho. \end{aligned}$$

It follows that

$$\int_0^{2\pi} |I_1|^2\, \tau^{-2\alpha-1} d\tau \leqq A \int_0^{2\pi} \tau^{-1-\alpha} d\tau \int_{\rho_T}^{1} |u'''_\alpha(\rho e^{i(\theta+\tau)})|^2 (1-\rho)^{5-\alpha} d\rho.$$

If we define $h_\rho(\tau)$ as 1 for $4\pi(1-\rho) \leqq \tau \leqq 2\pi$ and as 0 for $0 \leqq \tau < 4\pi(1-\rho)$, then

$$\int_0^{2\pi} |I_1|^2\, \tau^{-2\alpha-1} d\tau \leqq A \int_{\frac{1}{2}}^{1} (1-\rho)^{5-\alpha} d\rho \int_0^{2\pi} |u'''_\alpha(\rho e^{i(\theta+\tau)})|^2\, h_\rho(\tau)\tau^{-\alpha-1} d\tau.$$

We have

$$\begin{aligned} & u'''_\alpha(\rho e^{i(\theta+\tau)}) \\ & \quad = \frac{1}{2\pi} \int_0^{2\pi} u'(\rho^{\frac{1}{2}} e^{is}) \Big[\sum_1^\infty \frac{(n-1)(n-2)}{(in)^\alpha}\, \rho^{\frac{1}{2}n-5/2} e^{-2is} e^{i(n-3)(\theta+\tau-s)}\Big] ds. \end{aligned}$$

If $\rho \geqq \frac{1}{2}$, then $\mid \sum_{1}^{\infty}(n-1)(n-2)(in)^{-\alpha}\rho^{\frac{1}{2}n-5/2}e^{i(n-3)t} \mid \leqq A \mid 1-\rho e^{it} \mid^{-3+\alpha}$, and hence

$$\mid u'''_{\alpha}(\rho e^{i(\theta+\tau)}) \mid \leqq A \int_{0}^{2\pi} u'(\rho^{\frac{1}{2}}e^{is}) \mid \mid 1-\rho e^{i(\tau+\theta-s)} \mid^{-3+\alpha} ds,$$

(2)

$$\mid u'''_{\alpha}(\rho e^{i(\theta+\tau)}) \mid^{2} \leqq A(1-\rho)^{-3+2\alpha} \int_{0}^{2\pi} \mid u'(\rho^{\frac{1}{2}}e^{is}) \mid^{2} \mid 1-\rho e^{i(\tau+\theta-s)} \mid^{-2} ds.$$

We have

$$\int_{0}^{2\pi} \mid I_{1} \mid^{2} \tau^{-2\alpha-1}\, d\tau$$

$$\leqq A \int_{\frac{1}{2}}^{1} (1-\rho)^{2+\alpha} d\rho \int_{0}^{2\pi} \mid u'(\rho^{\frac{1}{2}}e^{is}) \mid^{2} ds \int_{0}^{2\pi} \mid 1-\rho e^{i(\tau+\theta-s)} \mid^{-2} h_{\rho}(\tau)\tau^{-\alpha-1}\, d\tau.$$

It is easily verified that

$$\int_{0}^{2\pi} \mid 1-\rho e^{i(\tau-\phi)} \mid h_{\rho}(\tau)\tau^{-\alpha-1} d\tau \leqq (1-\rho)^{-1} \mid \phi \mid^{-1-\alpha}, \quad 1-\rho \leqq \mid \phi \mid \leqq \pi,$$
$$\leqq (1-\rho)^{-2-\alpha}, \qquad 0 \leqq \mid \phi \mid \leqq 1-\rho.$$

Thus

$$\int_{0}^{2\pi} \mid I_{1} \mid^{2} \tau^{-2\alpha-1} d\tau \leqq A \int_{\frac{1}{2}}^{1} d\rho \int_{\mid \theta-s \mid \leqq 1-\rho} \mid u'(\rho^{\frac{1}{2}}e^{is}) \mid^{2} ds$$
$$+ A \int_{\frac{1}{2}}^{1} (1-\rho)^{1+\alpha} d\rho \int_{\mid \theta-s \mid \geqq 1-\rho} \mid u'(\rho^{\frac{1}{2}}e^{is}) \mid^{2} \mid \theta - s \mid^{-\alpha-1} ds.$$

Setting $\rho^{\frac{1}{2}} = r$, we obtain

$$\int_{0}^{2\pi} \mid I_{1} \mid^{2} \tau^{-2\alpha-1} d\tau \leqq A \int_{0}^{1} r dr \int_{\mid \theta-s \mid \leqq 2(1-r)} \mid u'(re^{is}) \mid^{2} ds$$
$$+ A \int_{0}^{1} (1-r)^{1+\alpha} dr \int_{\mid \theta-s \mid \geqq (1-r)} \mid u'(re^{is}) \mid^{2} \mid \theta - s \mid^{-\alpha-1} ds$$

$\leqq AG^{2}_{1+\alpha}(\theta) + AS^{2}(\theta)$. Using Theorems 1.2c and 1.2d, we find that

$$\int_{0}^{2\pi} [\int_{0}^{2\pi} \mid I_{1} \mid^{2} \tau^{-2\alpha-1}\, d\tau]^{\frac{1}{2}p} d\theta \leqq A \parallel u \parallel_{p}^{p}.$$

Consider I_{2}. We have

$$\mid I_{2} \mid \leqq A\tau^{2} \int_{\theta}^{\theta+\tau} \mid u'''_{\alpha}(\rho_{\tau}e^{it}) \mid dt, \qquad \mid I_{2} \mid^{2} \leqq A\tau^{5} \int_{\theta}^{\theta+\tau} \mid u'''_{\alpha}(\rho_{\tau}e^{it}) \mid^{2} dt.$$

Making use of (2) with $\rho = \rho_{\tau}$, we find that

$$| I_2 |^2 \leqq A\tau^{2+2\alpha} \int_\theta^{\theta+\tau} dt \int_0^{2\pi} | u'(\rho_\tau^{\frac{1}{2}} e^{is}) |^2 \, | 1 - \rho_\tau^{\frac{1}{2}} e^{i(t-s)} |^{-2} \, ds$$

$$\leqq A\tau^{2+2\alpha} \int_0^{2\pi} | u'(\rho_\tau^{\frac{1}{2}} e^{is}) |^2 \, ds \int_\theta^{\theta+\tau} | 1 - \rho_\tau^{\frac{1}{2}} e^{i(t-s)} |^{-2} \, dt.$$

Since $\int_\theta^{\theta+\tau} | 1 - \rho_\tau^{\frac{1}{2}} e^{i(t-s)} |^{-2} \, dt \leqq AP(\rho_\tau^{\frac{1}{2}}, \theta - s)$, we have

$$| I_2 |^2 \leqq A\tau^{2+2\alpha} \int_0^{2\pi} | u'(\rho_\tau^{\frac{1}{2}} e^{is}) |^2 \, P(\rho_\tau^{\frac{1}{2}}, \theta - s) ds,$$

$$\int_0^{2\pi} \tau^{-2\alpha-1} \, | I_2 |^2 d\tau \leqq A \int_0^{2\pi} d\tau \int_0^{2\pi} | u'(\rho_\tau^{\frac{1}{2}} e^{is}) |^2 \, P(\rho_\tau^{\frac{1}{2}}, \theta - s) ds$$

$$\leqq A \int_{\frac{1}{2}}^{1} (1 - \rho) d\rho \int_0^{2\pi} | u'(\rho^{\frac{1}{2}}, e^{is}) |^2 \, P(\rho^{\frac{1}{2}}, \theta - s) ds$$

$$\leqq A \int_0^{1} (1 - r) dr \int_0^{2\pi} | u'(re^{is}) |^2 \, P(r, \theta - s) ds.$$

From this it follows, as before, that

$$\int_0^{2\pi} [\int_0^{2\pi} \tau^{-2\alpha-1} \, | I_2 |^2 d\tau]^{\frac{1}{2}p} d\tau \leqq A \, \| u \|_p^p.$$

Consider I_3. We have $| I_3 | \leqq A \, | \tau |^2 \, | u''_\alpha(\rho_\tau e^{i\theta}) |$. Since

$$u''_\alpha(\rho^{i\theta}) = (\rho e^{i\theta})^{-2} [- u_{\alpha-2}(\rho e^{i\theta}) + i u_{\alpha-1}(\rho e^{i\theta})],$$

it follows that

$$\int_0^{2\pi} \tau^{-2\alpha-1} \, | I_3 |^2 d\tau \leqq A \int_{\frac{1}{2}}^{1} (1 - \rho)^{3-2\alpha} \, | u_{\alpha-2}(\rho e^{i\theta}) |^2 \, d\rho$$

$$+ A \int_{\frac{1}{2}}^{1} (1 - \rho)^{3-2\alpha} \, | u_{\alpha-1}(\rho e^{i\theta}) |^2 \, d\rho$$

$$\leqq A \int_0^{1} (1 - \rho)^{3-2\alpha} \, | u_{\alpha-2}(\rho e^{i\theta}) |^2 \, d\rho$$

$$+ A \int_0^{1} (1 - \rho)^{1-2\alpha} \, | u_{\alpha-1}(\rho e^{i\theta}) |^2 \, d\rho$$

$$\leqq A\Delta(\alpha - 1, u, \theta)^2 + A\Delta(\alpha, u, \theta)^2.$$

By Theorem 2.1, $\int_0^{2\pi} [\int_0^{2\pi} \tau^{-2\alpha-1} \, | I_3 |^2 \, d\tau]^{\frac{1}{2}p} d\theta \leqq A \, \| u \|_p^p$.

Similar arguments serve to show that

$$\int_0^{2\pi} [\int_0^{2\pi} \tau^{-2\alpha-1} \, | I_k |^2 d\tau]^{\frac{1}{2}p} \, d\theta \leqq A \, \| u \|_p^p, \qquad k = 4, 5, 6.$$

Combining these six inequalities, we obtain our theorem.

3.2. We shall here prove the converse inequality.

THEOREM 3.2. *If* $1 < p < \infty$, $0 < \alpha < 1$, *then* $\| \delta(\alpha, u, \theta) \|_p \geqq A \| u \|_p$.

We will show that

$$\delta(\alpha, u, \theta) > A \Delta(\alpha, u, \theta). \tag{1}$$

Our result will then follow from Theorem 2.2. We have

$$u_\alpha(\rho, \theta) = \int_0^{2\pi} u_\alpha(t) P(\rho, \theta - t)\, dt.$$

Differentiating with respect to θ,

$$u_{\alpha-1}(\rho, \theta) = \int_0^{2\pi} u_\alpha(t) P_\theta(\rho, \theta - t)\, dt.$$

Since P_θ is odd, we find that

$$u_{\alpha-1}(\rho, \theta) = \tfrac{1}{2} \int_0^{2\pi} [u_\alpha(\theta - \tau) - u_\alpha(\theta + \tau)] P_\theta(\rho, \tau)\, d\tau.$$

Now $| P_\theta(\rho, \tau) | \leqq A | 1 - \rho e^{i\tau} |^{-2}$. By Schwarz's inequality

$$| u_{\alpha-1}(\rho, \theta) | \leqq A \int_0^{2\pi} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) | \, | 1 - \rho e^{i\tau} |^{-2}\, d\tau,$$

$$\begin{aligned} | u_{\alpha-1}(\rho, \theta) |^2 &\leqq A \int_0^{2\pi} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) |^2 \, | 1 - \rho e^{i\tau} |^{-2-\alpha}\, d\tau \\ &\qquad \times \int_0^{2\pi} | 1 - \rho e^{i\tau} |^{-2+\alpha}\, d\tau \\ &\leqq A (1 - \rho)^{-1+\alpha} \int_0^{2\pi} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) |^2 \, | 1 - \rho e^{i\tau} |^{-2-\alpha}\, d\tau. \end{aligned}$$

Thus

$$\begin{aligned} \int_0^1 (1 - \rho)^{1-2\alpha} | u_{\alpha-1}(\rho, \theta) |^2\, d\rho &\leqq A \int_0^1 (1 - \rho)^{-\alpha}\, d\rho \int_0^{2\pi} | u_\alpha(\theta + \tau) \\ &\qquad - u_\alpha(\theta - \tau) |^2 \, | 1 - \rho e^{i\tau} |^{-2-\alpha}\, d\tau, \\ &\leqq A \int_0^{2\pi} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) |^2\, d\tau \int_0^1 (1 - \rho)^{-\alpha} | 1 - \rho e^{i\tau} |^{-2-\alpha}\, d\rho. \end{aligned}$$

Now $\int_0^1 (1 - \rho)^{-\alpha} | 1 - \rho e^{i\tau} |^{-2-\alpha}\, d\rho \leqq A \tau^{-1-2\alpha}$, and thus

$$\int_0^1 (1 - \rho)^{1-2\alpha} | u_{\alpha-1}(\rho, \theta) |^2\, d\rho \leqq A \int_0^{2\pi} | u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) |^2 \tau^{-2\alpha-1}\, d\tau.$$

Theorems 3. 1 and 3. 2 are false if $\alpha = 1$. However, they can be modified so that they become true. Marcenkiewicz [6] conjectured and Zygmund [8] proved that

$$A \| u \|_p^p \leqq \int_0^{2\pi} [\int_0^{2\pi} | u_1(\theta + \tau) - 2u_1(\theta) + u_1(\theta - \tau)|^2 \tau^{-3} \, d\tau]^{\frac{1}{2}p} \, d\theta$$
$$\leqq A \| u \|_p^p.$$

4. 1. Let $u(\theta) \in L_1(0, 2\pi)$ and have mean value zero. Let $u_{\alpha-1}(\rho, \theta)$ be defined as in § 1. 1 and let $u\dagger(\theta)$ be defined as in § 1. 2.

LEMMA 4. 1. $\qquad | u_{\alpha-1}(\rho, \theta) | \leqq A(1-\rho)^{-1+\alpha} u\dagger(\theta).$

We have $u_{\alpha-1}(\rho, \theta) = \int_{-\pi}^{\pi} u(\theta - t) k(\alpha, \rho, t) \, dt$, where

$$k(\alpha, \rho, t) = \tfrac{1}{2\pi} \Sigma' e^{int} \rho^{|n|} / (in)^{\alpha-1}.$$

Let $U(\theta, t) = \int_0^t u(\theta - \tau) \, d\tau$. We have $U(\theta, t) \leqq | t | \, u\dagger(\theta)$. Integrating by parts we obtain, since the integrated term vanishes,

$$u_{\alpha-1}(\rho, \theta) = - \int_{-\pi}^{\pi} U(\theta, t) \partial k(\alpha, \rho, t)/\partial t \, dt,$$

$$| u_{\alpha-1}(\rho, \theta) | \leqq u\dagger(\theta) \int_{-\pi}^{\pi} | t \, \partial k(\alpha, \rho, t)/\partial t | \, dt.$$

Now $\partial k(\alpha, \theta, t)/\partial t \, | \leqq A \, | 1 - \rho e^{it} |^{-3+\alpha}$, from which it follows that

$$\int_{-\pi}^{\pi} | t \, \partial k(\alpha, \rho, t)/\partial t \, | \, dt \leqq A(1-\rho)^{-1+\alpha}$$

Inserting this, our lemma is proved.

Let $1 < r < \infty$. We define

$$\Delta(r, \alpha, u, \theta) = [\int_0^1 | u_{\alpha-1}(\rho, \theta) |^r \, (1-\rho)^{r-1-\alpha r} \, d\rho]^{1/r}.$$

THEOREM 4. 1. *If* $1 < p < \infty$, $-\infty < \alpha < 1$, *then*

$$\| \Delta(r, \alpha, u, \theta) \|_p \leqq A \, \| u \|_p \text{ for } 2 \leqq r < \infty;$$

$$\| \Delta(r, \alpha, u, \theta) \|_p \geqq A \, \| u \|_p \text{ for } 1 < r \leqq 2.$$

Suppose first that $r \geqq 2$. Then, by Lemma 4. 1,

$$\Delta(r, \alpha, u, \theta)^r = \int_0^1 | u_{\alpha-1}(\rho, \theta) |^r (1-\rho)^{r-1-\alpha r} d\rho$$

$$\leqq A[\int_0^1 | u_{\alpha-1}(\rho, \theta) |^2 (1-\rho)^{1-2\alpha} d\rho][u\dagger(\theta)]^{r-2},$$

so that $\Delta(r, \alpha, u, \theta) \leqq A\Delta(2, \alpha, u, \theta)^{2/r}[u\dagger(\theta)]^{1-2/r}$. Using Hölder's inequality and Theorem 2. 1, we obtain

$$\| \Delta(r, \alpha, u, \theta) \|_p \leqq A \| \Delta(2, \alpha, u, \theta) \|_p^{2/r} \| u\dagger(\theta) \|_p^{1-2/r} \leqq A \| u \|_p^p.$$

Next let $1 < r \leqq 2$. We have

$$\Delta(2, \alpha, u, \theta)^2 = \int_0^1 | u_{\alpha-1}(\rho, \theta) |^2 (1-\rho)^{1-2\alpha} d\rho$$

$$\leqq A[\int_0^1 | u_{\alpha-1}(\rho, \theta) |^r (1-\rho)^{r-1-\alpha r} d\rho][u\dagger(\theta)]^{2-r},$$

so that $\Delta(2, \alpha, u, \theta) \leqq A\Delta(r, \alpha, u, \theta)^{\frac{1}{2}r}[u\dagger(\theta)]^{1-\frac{1}{2}r}$. Using Hölder's inequality and Theorem 2. 2, we obtain

$$\| u \|_p \leqq A \| \Delta(r, \alpha, u, \theta) \|_p^{\frac{1}{2}r} \| u \|_p^{1-\frac{1}{2}r}, \qquad \| \Delta(r, \alpha, u, \theta) \|_p \geqq A \| u \|_p.$$

For similar results see [7].

4. 2. The two lemmas which follow are analogous to results proved in Hardy and Littlewood [2].

LEMMA 4. 2a. *If* $u(\theta) \in L^p(0, 2\pi)$, $1 < p < \infty$, $0 < \alpha < 1$, *and* $1/p + 1/q = 1$, *then*

$$\int_0^{2\pi} \| u(\theta+\tau) - u(\theta-\tau) \|_p^p \tau^{-1-p\alpha} d\tau$$

$$\leqq A \int_0^1 \| u_{-1}(\rho, \theta) \|_p^p (1-\rho)^{p(1/q-\alpha)} d\rho.$$

Let $\rho_\tau = 1 - (\tau/4\pi)$. We have $u(\theta+\tau) - u(\theta-\tau) = I_1 + I_2 + I_3$, where

$$I_1 = \int_{\rho_\tau}^1 \partial u(\rho, \theta+\tau)/\partial\rho d\rho, \qquad I_2 = \int_{-\tau}^{\tau} \partial u(\rho_\tau, \theta+\phi)/\partial\theta d\phi,$$

$$I_3 = \int_1^{\rho_\tau} \partial u(\rho, \theta-\tau)/\partial\rho d\rho.$$

By Minkowski's inequality,

$$\| I_2 \|_p = \left(\int_0^{2\pi} d\theta \left| \int_{-\tau}^{\tau} \partial u(\rho_\tau, \theta + \phi)/\partial\theta d\phi \right|^p \right)^{1/p}$$
$$\leqq \int_{-\tau}^{\tau} d\phi \left(\int_0^{2\pi} | u_{-1}(\rho_\tau, \theta + \phi) |^p d\theta \right)^{1/p}$$

$\leqq A \mid \tau \mid \| u_{-1}(\rho_\tau, \theta) \|_p$. Thus

$$(1) \qquad \int_0^{2\pi} \| I_2 \|_p^p \tau^{-1-p\alpha} d\tau \leqq A \int_{\frac{1}{2}}^1 \| u_{-1}(\rho, \theta) \|_p^p (1 - \rho)^{p-1-p\alpha} d\rho$$

$\leqq A \int_0^1 \| u_{-1}(\rho, \theta) \|_p^p (1 - \rho)^{p-1-p\alpha} d\rho$. Consider I_2. Since $\partial u/\partial\rho = \tilde{u}_{-1}/\rho$,

$$\| I_1 \|_p = \left(\int_0^{2\pi} d\theta \left| \int_{\rho_\tau}^1 1/\rho \tilde{u}_{-1}(\rho, \theta + \tau) d\rho \right|^p \right)^{1/p}$$
$$\leqq \int_{\rho_\tau}^1 1/\rho \| \tilde{u}_{-1}(\rho, \theta) \|_p d\rho \leqq A \int_{\rho_\tau}^1 \| u_{-1}(\rho, \theta) \|_p d\rho.$$

Here we have used the fact that $\rho \geqq \frac{1}{2}$ for all τ and M. Riesz's theorem on conjugate functions. Let λ be a parameter such that $q\lambda < 1$, $p\lambda > -1$. We have

$$\| I_1 \|_p \leqq \int_{\rho_\tau}^1 \| u_{-1}(\rho, \theta) \|_p (1 - \rho)^{\lambda} (1 - \rho)^{-\lambda} d\rho.$$

By Hölder's inequality,

$$\| I_1 \|_p^p \leqq \int_{\rho_\tau}^1 \| u_{-1}(\rho, \theta) \|_p^p (1 - \rho)^{p\lambda} d\rho \left(\int_{\rho_\tau}^1 (1 - \rho)^{-q\lambda} d\rho \right)^{p/q}.$$

We have

$$\int_{\rho_\tau}^1 (1 - \rho)^{-q\lambda} d\rho \leqq A (1 - \rho_\tau)^{1-q\lambda},$$

$$\| I_1 \|_p^p \tau^{-\alpha p-1} \leqq A (1 - \rho_\tau)^{-\lambda p-\alpha p-1+p/q} \int_{\rho_\tau}^1 \| u_{-1}(\rho, \theta) \|_p^p (1 - \rho)^{p\lambda} d\rho,$$

$$\int_0^{2\pi} \| I_1 \|_p^p \tau^{-\alpha p-1} d\tau$$
$$\leqq A \int_0^{2\pi} (1 - \rho_\tau)^{-\lambda p-\alpha p-1+p/q} d\tau \int_{\rho_\tau}^1 \| u_{-1}(\rho, \theta) \|_p^p (1 - \rho)^{p\lambda} d\rho$$
$$\leqq A \int_{\frac{1}{2}}^1 \| u_{-1}(\rho, \theta) \|_p^p (1 - \rho)^{p\lambda} d\rho \int_{4\pi(1-\rho)}^{2\pi} (1 - \rho_\tau)^{-\lambda p-\alpha p-1+p/q} d\tau.$$

If $-\lambda p - \alpha p + p/q < 0$, which for $\alpha > 0$ can be effected by a suitable choice of λ, then

$$\int_{4\pi(1-\rho)}^{2\pi} (1-\rho_\tau)^{-\lambda p-\alpha p-1+p/q} d\tau \leqq A(1-\rho)^{-\lambda p-\alpha p+p/q},$$

$$(2) \qquad \int_0^{2\pi} \| I_1 \|_p^p \tau^{-\alpha p-1} d\tau \leqq A \int_0^1 \| u_{-1}(\rho,\theta) \|_p^p (1-\rho)^{-\alpha p+p/q} d\rho.$$

In an exactly similar way we can show that

$$(3) \qquad \int_0^{2\pi} \| I_3 \|_p^p \tau^{-\alpha p-1} d\tau \leqq A \int_0^1 \| u_{-1}(\rho,\theta) \|_p^p (1-\rho)^{-\alpha p+p/q} d\rho.$$

By Jensen's inequality,

$$\| u(\theta+\tau) - u(\theta-\tau) \|_p^p \leqq 3^{p-1}(\| I_1 \|_p^p + \| I_2 \|_p^p + \| I_3 \|_p^p)$$

and thus, using (1), (2) and (3),

$$\int_0^{2\pi} \| u(\theta+\tau) - u(\theta-\tau) \|_p^p \tau^{-1-\alpha p} d\tau \leqq A \int_0^1 \| u_{-1}(\rho,\theta) \|_p^p (1-\rho)^{-\alpha p+p/q} d\rho.$$

LEMMA 4.2b. *If* $1 < p < \infty$, $-1/p < \alpha < 1$, $1/p + 1/q = 1$ *if*

$$u(\theta) \in L^p(0, 2\pi),$$

then

$$\int_0^1 \| u_{-1}(\rho,\theta) \|_p^p (1-\rho)^{p(1/q-\alpha)} d\rho$$

$$\leqq A \int_0^1 \| u(\theta+\tau) - u(\theta-\tau) \|_p^p \tau^{-1-\alpha p} d\tau.$$

We have

$$u_{-1}(\rho,\theta) = \tfrac{1}{2} \int_0^{2\pi} [u(\theta-\tau) - u(\theta+\tau)]\, \partial P(\rho,\tau)/\partial\tau d\tau.$$

Since $| \partial/P(\rho,\tau)\partial\tau | \leqq A \,| \rho e^{i\tau} - 1 |^{-2}$, we have, using Hölder's inequality,

$| u_{-1}(\rho,\theta)|^p$

$$\leqq A \int_0^{2\pi} | u(\theta+\tau) - u(\theta-\tau)|^p \,| \rho e^{i\tau} - 1 |^{-p\lambda} d\tau \Big[\int_0^{2\pi} | \rho e^{i\tau} - 1 |^{-q\mu} d\tau \Big]^{p/q}.$$

Here λ and μ are parameters such that $\lambda + \mu = 2$, $q\mu > 1$. Because of this last condition $\int_0^{2\pi} | \rho e^{i\tau} - 1 |^{-q\mu} d\tau \leqq A(1-\rho)^{1-q\mu}$, and thus

$$| u_{-1}(\rho,\theta)|^p \leqq A(1-\rho)^{-p\mu+p/q} \int_0^{2\pi} | u(\theta+\tau) - u(\theta-\tau)|^p \,| e^{i\tau}\rho - 1 |^{-p\lambda} d\tau,$$

$$\| u_{-1}(\rho,\theta)\|_p^p \leqq A(1-\rho)^{-p\mu+p/q} \int_0^{2\pi} \| u(\theta+\tau) - u(\theta-\tau)\|_p^p \,| e^{i\tau}\rho - 1 |^{-p\lambda} d\tau,$$

$$\int_0^1 \| u_{-1}(\rho,\theta) \|_p^p (1-\rho)^{-p\alpha+p/q} d\rho$$

$$\leqq A \int_0^{2\pi} \| u(\theta+\tau) - u(\theta-\tau) \|_p^p \, d\tau \int_0^1 (1-\rho)^{-p\alpha - p\mu + 2p/q} \, | e^{i\tau}\rho - 1 |^{-p\lambda} d\rho.$$

If $-p\alpha - p\mu + 2p/q > -1$, $p\lambda + p\mu + p\alpha - 2p/q > 1$ then

$$\int_0^1 (1-\rho)^{-p\alpha - p\mu + 2p/q} \, | e^{i\tau}\rho - 1 |^{-p\lambda} d\rho \leqq A \, | \tau |^{-1-p\alpha}$$

and

$$\int_0^1 \| u_{-1}(\rho, \theta) \|_p^p \, (1-\rho)^{-p\alpha + p/q} d\rho$$

$$\leqq A \int_0^{2\pi} \| u(\theta+\tau) - u(\theta-\tau) \|_p^p \tau^{-1-p\alpha} d\tau.$$

If $-1/p < \alpha < 1$, then the conditions on the parameters can be satisfied.

Using these we can prove a result which is analogous to Theorem. 4. 1 for the special choice of the parameter r, $r = p$.

THEOREM 4. 2a. *If* $0 < \alpha < 1$, *then*

$$\int_0^{2\pi} \| u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) \|_p^p \tau^{-1-p\alpha} d\tau \geqq A \, \| u \|_p^p, \qquad 1 < p \leqq 2;$$

$$\int_0^{2\pi} \| u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) \|_q^q \tau^{-1-q\alpha} d\tau \leqq A \, \| u \|_q^q, \qquad 2 \leqq q < \infty.$$

This is a consequence of Lemmas 4. 2a and 4. 2b and Theorem 4. 1. The following result complements Theorem 4. 2a.

THEOREM 4. 2b. *If* $0 < \alpha < 1$, *then*

$$\int_0^{2\pi} \| u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) \|_p^2 \, \tau^{-1-2\alpha} d\tau \geqq A \, \| u \|_p^2, \qquad 1 < p \leqq 2;$$

$$\int_0^{2\pi} \| u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) \|_q^2 \, \tau^{-1-2\alpha} d\tau \leqq A \, \| u \|_q^2, \qquad 2 \leqq q < \infty.$$

Minkowski's inequality gives

$$\int_0^{2\pi} \| u_\alpha(\theta+\tau) - u_\alpha(\theta-\tau) \|_r^2 \, \tau^{1-2\alpha} d\tau \geqq [\int_0^{2\pi} \delta(\alpha, u, \theta)^r d\theta]^{2/r} \quad (r \geqq 2),$$

$$\leqq [\int_0^{2\pi} \delta(\alpha, u, \theta)^r d\theta]^{2/r} \quad (r \leqq 2).$$

Appealing to Theorems 3. 1 and 3. 2 we obtain our desired result.

The result given in the introduction is an immediate corollary of these two theorems combined with the relation

$$\| u_\alpha(\theta + \tau) - u_\alpha(\theta - \tau) \|_p \leqq A \| u \|_p \, | \tau |^\alpha.$$

WASHINGTON UNIVERSITY AND
THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] G. H. Hardy and J. E. Littlewood, "Some properties of fractional integrals, I," *Mathematische Zeitschrift*, vol. 27 (1927), pp. 565-606.

[2] ———, "A convergence criterion for Fourier series," *ibid.*, vol. 28 (1928), pp. 612-634.

[3] ———, "Some properties of fractional integrals, II," *ibid.*, vol. 34 (1932), pp. 403-439.

[4] ———, "Theorems concerning mean values of analytic or harmonic functions," *Quarterly Journal of Mathematics, Oxford Series*, vol. 12 (1941), pp. 221-256.

[5] J. E. Littlewood and R. E. A. C. Paley, "Theorems on Fourier series and power series II," *Proceedings of the London Mathematical Society*, vol. 42 (1936), pp. 52-89.

[6] J. Marcinkiewicz, "Sur quelques integrales du type de Dini," *Annales de la Société Polonaise de Mathématique*, vol. 17 (1938), pp. 42-50.

[7] ——— and A. Zygmund, "A theorem of Lusin," *Duke Mathematical Journal*, vol. 4 (1938), pp. 473-485.

[8] A. Zygmund, "On certain integrals," *Transactions of the American Mathematical Society*, vol. 55 (1944), pp. 170-204.

[9] ———, "Proof of a theorem of Littlewood and Paley," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 439-446.

[10] ———, *Trigonometrical Series*, Warsaw-Lwow, 1935.

A CLASS OF d-SIMPLE SEMIGROUPS.*

By A. H. CLIFFORD.

In a recent paper, J. A. Green [2] proposed the determination of all regular d-simple semigroups (definitions in § 1 below) as the next step beyond the determination by Suschkewitsch [4] and Rees [3] of all completely simple semigroups. The present paper takes a small portion of this step, dealing with semigroups S satisfying the following conditions:

A1. *S is d-simple.*

A2. *S has an identity element.*

A3. *Any two idempotent elements of S commute.*

(Regularity is a consequence of A1 and A2.) It is shown that the structure of S is determined by that of its right unit subsemigroup P, and that P has the following properties:

B1. *The right cancellation law holds in P.*

B2. *P has an identity element.*

B3. *The intersection of two principal left ideals of P is a principal left ideal.*

Conversely, if P is any semigroup having properties B1, 2, 3, there exists a semigroup S satisfying A1, 2, 3, the right unit subsemigroup of which is isomorphic with P. A detailed statement of this result is given in § 1, and the rest of the paper is devoted to the proof thereof.

The problem of describing all semigroups S satisfying A1, 2, 3 is of course merely replaced by the perhaps equally difficult one of describing all semigroups P satisfying B1, 2, 3. The extent of the class of all such P can be seen from the fact that the positive part of a lattice-ordered group ([1], Chapter 14) satisfies not only B1, 2, 3 but also the left-right duals thereof.

The present work parallels rather than extends that of Suschkewitsch and Rees. For while a completely simple semigroup is regular and d-simple, it reduces to a group if either of the conditions A2 or A3 is imposed upon it.

* Received November 26, 1952.

1. Definitions and statement of main result. Let S be any semigroup with identity element 1. We say that two elements a and b of S are right associate if they generate the same principal right ideal: $aS = bS$. An element p of S is right associate to 1 if and only if it has a right inverse q in S: $pq = 1$. Such an element p will be called a right unit of S. The set P of all right units of S is a subsemigroup of S which we shall call the right unit subsemigroup of S. P evidently satisfies B1 and B2. Left associate elements, left units, and the left unit subsemigroup Q of S, are defined analogously. Two elements of S will be called associate if they are both left and right associate. The intersection $U = P \cap Q$ of P and Q consists of all elements associate to 1, i. e. of all elements having inverses on both sides. U is evidently a subgroup of S. Its elements will be called units.

Two elements a and b of S are called d-equivalent (Green [2], p. 164) if there exists an element S which is left associate to a and right associate to b. (This implies the existence of an element of S which is right associate to a and left associate to b.) We shall say that S is d-simple if it consists of a single class of d-equivalent elements. A d-simple semigroup is simple, but not conversely.

Now let P be any semigroup satisfying B1, 2, 3. From each class of left associate elements of P let us pick a fixed representative. B3 states that if a and b are elements of P, there exists c in P such that $Pa \cap Pb = Pc$. c is determined by a and b only to within left associates. We define $a \vee b$ to be the representative of the class to which c belongs. $a \vee b$ evidently has the properties of a least common left multiple (LCLM) of a and b. We observe also that $a \vee b = b \vee a$. We define a binary operation $*$ by

$$(1.1) \qquad (a * b)b = a \vee b$$

for each pair of elements a, b of P. There is at least one such element in P since $a \vee b$ is in Pb, and at most one by B1.

Now let $P^{-1} \circ P$ denote the set of ordered pairs (a, b) of elements of P with equality defined by

$$(1.2) \qquad (a, b) = (a', b') \text{ if } a' = ua, b' = ub$$

where u is a unit in P. We define product in $P^{-1} \circ P$ by

$$(1.3) \qquad (a, b)(c, d) = ((c * b)a, (b * c)d).$$

MAIN THEOREM. *Starting with a semigroup P satisfying* B1, 2, 3, *equations* (1.1), (1.2), (1.3) *define a semigroup $P^{-1} \circ P$ satisfying* A1, 2, 3; *P is isomorphic with the right unit subsemigroup of $P^{-1} \circ P$. Conversely, if*

S is a semigroup satisfying A1, 2, 3, *its right unit subsemigroup P satisfies* B1, 2, 3 *and S is isomorphic with $P^{-1} \circ P$.*

Of course we shall have to show that (1. 3) is single-valued, i. e. consistent with (1. 2), and that the associative law holds, in addition to the properties mentioned. But let us point out here that *the definition* (1. 3) *is independent of the choice of representative elements in the classes of left associates in P.* For a new choice would (by Lemma 4. 1) replace each $a \vee b$ by $u_{a,b}$ $(a \vee b)$ with $u_{a,b}$ a unit. From $(a * b)b = a \vee b = b \vee a = (b * a)a$ we see that $a * b$ and $b * a$ are multiplied on the left by the same $u_{a,b}$. By (1. 2), this has no effect on the product as defined by (1. 3).

Let us also mention here without proof that, in the case $U = \{1\}$, $P^{-1} \circ P$ can be isomorphically represented by the semigroup of mappings of P into itself generated by the mappings p_a, q_a (a an arbitrary element of P) defined as follows: $q_a x = xa$, $p_a x = x * a$ (x a variable element of P). One shows that $q_a q_b = q_{ba}$, $p_a p_b = p_{ab}$, and $p_a q_b = q_{b*a} p_{a*b}$. This construction avoids the troublesome proof of associativity, but does not seem applicable in the presence of units $\neq 1$.

2. d-simple semigroups with identity element. In this section we deal with semigroups S satisfying A1 and A2, but not necessarily A3.

THEOREM 2. 1. *Let S be a semigroup with identity element, and let P and Q be its right and left unit subsemigroups. Then S is d-simple if and only if $S = QP$.*

Proof. Assume first that S is d-simple, and let a be an arbitrary element of S. Since a is d-equivalent to 1, there exists an element p of S which is left associate to a and right associate to 1. The latter means that p is in P. The former means that elements x and y exist in S such that $a = xp$ and $p = ya$. Let q be a right inverse of p. Then $1 = pq = yaq = yxpq = yx$. Hence x is in Q, and from $a = xp$ we infer $S = QP$.

Suppose conversely that $S = QP$. Let $a = qp (p \in P, q \in Q)$ be any element of S. By definition of P and Q there exist x and y in S such that $px = yq = 1$. From $a = qp$ and $p = ya$ we see that a and p are left associate. Since 1 and p are right associate, a is d-equivalent to 1, and S is d-simple.

A semigroup S is said to be regular if to each a in S there exists an element x of S such that $axa = a$ (Green [2], p. 163).

COROLLARY. *A d-simple semigroup S with identity element is regular.*

Proof. Let $a = qp\,(p \varepsilon P, q \varepsilon Q)$ be any element of S. There exist y and z in S such that $py = zq = 1$. Let $x = yz$. Then

$$axa = qpyzqp = q \cdot 1 \cdot 1 \cdot p = qp = a.$$

In Lemmas 2.1-2.4, S will denote a semigroup with identity element, P and Q its right and left unit subsemigroups, and $U = P \cap Q$ its group of units. Lemmas 2.2-2.4 are concerned with properties of the complex QP. Our ultimate interest is, of course, in the case $QP = S$, but the validity of these lemmas does not depend on this assumption.

LEMMA 2.1. *The group of units of P is the group U of units of S. For any element p of P, $Pp = P \cap Sp$, i. e. an element p' of P is a left multiple of p in S if and only if it is a left multiple of p in P. Two elements of P are associate in S if and only if they are left associate in P. This in turn holds if and only if the two elements differ by a unit factor in the left; this unit factor is unique.*

Proof. $U = P \cap Q$, where Q is the left unit subsemigroup of S, and hence U is contained in the group of units of P. But a unit in P has a left inverse in P, hence also in S, hence belongs to Q, and therefore to $P \cap Q = U$. If p' is a left multiple of p in S, say $p' = xp$, and if q' is a right inverse of p' in S, then $1 = p'q' = x \cdot pq'$. Hence $x \varepsilon P$, and p' is a left multiple of p in P.

Consequently, two elements p and p' of P are left associate in P if and only if they are left associate in S, hence if and only if they are associate in S, since any two elements of P are right associate in S. If this is the case, then $p = xp'$ and $p' = yp$ with x and y in P. From $p = xyp$, $p' = yxp'$, and right cancellation in P, we infer $xy = yx = 1$, i. e. x and y are units. The unicity of x and y also follow from right cancellation in P.

The left-right dual of Lemma 2.1 also holds, the proof being similar.

LEMMA 2.2. *An element $e = qp$ of QP is idempotent if and only if $pq = 1$.*

Proof. If $pq = 1$, $e^2 = qpqp = q1p = qp = e$. Suppose conversely that $e = qp$ is idempotent. Since $p \varepsilon P$ and $q \varepsilon Q$, elements x and y exist in S such that $px = yq = 1$. Then $pq = yq \cdot pq \cdot px = y(qp)^2x = yqpx = 1 \cdot 1 = 1$.

LEMMA 2.3. *If $a = qp$ is any element of QP, then p [q] is left [right] associate to a. If $a = q'p'$ is any other expression of a as the product of an element q' of Q by an element p' of P, then there exists a unit u such that $p' = up$, $q' = qu^{-1}$.*

Proof. Let x [y] be a left [right] inverse of q [p] : $xq = py = 1$. From $a = qp$ and $p = xa$ we infer $Sa = Sp$. From $a = qp$ and $q = ay$ we infer $aS = qS$. If $a = q'p'$ then $Sp = Sa = Sp'$ and $qS = aS = q'S$. From Lemma 2. 1 there exist units u and v such that $p' = up$, $q' = qv$. Then

$$1 = xqpy = xay = xq'p'y = xqvupy = vu.$$

Hence $p' = up$, $q' = qu^{-1}$.

LEMMA 2. 4. *If $a = qp$ is any element of QP, then any element of QP associate to a is expressible in the form qup with u a unit, and u is unique.*

Proof. Let $b = q'p'$ be an element of QP associate to a. Then, by Lemma 2. 3, $Sp' = Sb = Sa = Sp$, and, by Lemma 2. 1, $p' = vp$ with v a unit. Similarly $q' = qw$ with w a unit. Then $u = wv$ is a unit, and $b = q'p' = qwvp = qup$. Suppose $b = qup = qu'p$ with units u and u'. There exist elements x, y of S such that $xq = py = 1$. Then

$$u = xq \cdot u \cdot py = xby = xq \cdot u' \cdot py = u'.$$

THEOREM 2. 2. *The following three conditions on a d-simple semigroup S with identity element are equivalent.*

A3. *The idempotent elements of S commute.*

A4. *Every principal left ideal of S, and every principal right ideal, has a unique idempotent generator.*

A5. *Every right unit of S has a unique right inverse, and every left unit a unique left inverse in S.*

Proof. I. A3 implies A4. That every principal left or right ideal of S has at least one idempotent generator follows from the regularity of S (Corollary to Theorem 2. 1). To show the unicity, suppose $eS = fS\,(e^2 = e,$ $f^2 = f)$. Then $e = fx$ and $f = ey$ for some x and y in S, whence $fe = e$ and $ef = f$. A3 then implies $e = f$.

II. A4 implies A5. Suppose q and q' are right inverses of the element p of P: $pq = pq' = 1$. Then, by Lemmas 2. 2 and 2. 3, $e = qp$ and $e' = q'p$ are idempotent elements generating the same principal left ideal Sp. By A4, $e = e'$. Then $q = qpq = eq = e'q = q'pq = q'$.

III. A5 implies A3. Let e_1 and e_2 be any two idempotent elements of S. Since S is d-simple, $S = QP$ by Theorem 2. 1, and hence $e_1 = q_1p_1$ and $e_2 = q_2p_2$ with q_1 and q_2 in Q, p_1 and p_2 in P. By Lemma 2. 2, $p_1q_1 = p_2q_2$ $= 1$. Again using $S = QP$, $p_1q_2 = q_3p_4$ and $p_2q_1 = q_5p_6$ with q_3 and q_5 in Q,

p_4 and p_6 in P. Let p_3 and p_5 be respective left inverses of q_3 and q_5; likewise q_4 and q_6 right inverses of p_4 and p_6. Then $p_3p_1q_2q_4 = p_3q_3p_4q_4 = 1$, while $p_3p_1q_1q_3 = 1$ and $p_4p_2q_2q_4 = 1$. By A5 we infer that $q_1q_3 = q_2q_4$ and $p_3p_1 = p_4p_2$. Similarly, from $p_5p_2q_1q_6 = p_5q_5p_6q_6 = 1$, $p_5p_2q_2q_5 = 1$, and $p_6p_1q_1q_6 = 1$, we infer that $q_1q_6 = q_2q_5$ and $p_6p_1 = p_5p_2$. Then

$$p_4q_5 \cdot p_6q_3 = p_4p_2q_1q_3 = p_3p_1q_1q_3 = 1 \quad \text{and} \quad p_6q_3 \cdot p_4q_5 = p_6p_1q_2q_5 = p_6p_1q_1q_6 = 1.$$

Hence $p_4q_5 = u$ and $p_6q_3 = u^{-1}$ with u a unit. From $p_4 \cdot q_5u^{-1} = 1$ and $up_6 \cdot q_3 = 1$ and A5 we conclude that $q_5u^{-1} = q_4$ and $up_6 = p_3$. Finally,

$$\begin{aligned} e_1e_2 = q_1p_1q_2p_2 = q_1q_3p_4p_2 = q_2q_4p_3p_1 &= q_2q_5u^{-1}up_6p_1 \\ &= q_2q_5p_6p_1 = q_2p_2q_1p_1 = e_2e_1. \end{aligned}$$

3. Introduction of condition A3. We now assume that S is a semigroup satisfying A1, 2, 3. Then S has the properties A4 and A5 of Theorem 2.2. By A5, the relation $pq = 1$ defines a one-to-one correspondence $p \leftrightarrow q$ between P and Q which is evidently an anti-isomorphism.

We shall now adopt a new notation. Elements of P will be denoted by $a, b, c, \cdots$, and their corresponding right inverses in Q by $a^{-1}, b^{-1}, c^{-1}, \cdots$. Thus we have the rules $aa^{-1} = 1$, $(ab)^{-1} = b^{-1}a^{-1}$. By Theorem 2.1, every element of S is expressible in the form $a^{-1}b$ with a and b in P. By Lemma 2.2, $a^{-1}a$ is an idempotent element of S which we shall denote by e_a. Thus $a^{-1}a = e_a$, $ae_a = a$, $e_aa^{-1} = a^{-1}$. Elements of U will, as before, be denoted by $u, v, w, \cdots$. $e_a = 1$ if and only if $a \in U$.

LEMMA 3.1. *$e_a = e_b$ if and only if $Pa = Pb$ $(a, b \in P)$.*

Proof. By Lemma 2.1, $Pa = Pb$ if and only if $Sa = Sb$. Since $Sa = Se_a$ and $Sb = Se_b$, $Pa = Pb$ if and only if $Se_a = Se_b$, and by A4 (Theorem 2.2), the latter holds if and only if $e_a = e_b$.

By A3, the product of two idempotents is idempotent. The set E of idempotents of S is thus a commutative subsemigroup of S, and hence a semilattice (cf. [1], Example 1, p. 18 and Example 4, p. 25) under the usual definition $e \leqq f$ if $ef = fe = e$. By Lemma 2.2, every idempotent of S has the form e_a with $a \in P$. For given $a, b \in P$, e_ae_b is idempotent, and hence there exists an element $c \in P$ such that e_ae_b $(= e_be_a) = e_c$.

LEMMA 3.2. *If $e_ae_b = e_be_a = e_c$ $(a, b, c \in P)$, then $Pa \cap Pb = Pc$, and conversely.*

Proof. Suppose $e_a e_b = e_b e_a = e_c$. Since $S = QP$ (Theorem 2.1), there exist a_1 and b_1 in P such that $ab^{-1} = b_1^{-1}a_1$. From

$$c^{-1}c = a^{-1}a \cdot b^{-1}b = a^{-1}b_1^{-1}a_1b = (b_1a)^{-1}(a_1b)$$

and Lemma 2.3 we infer that c is left associate to a_1b: $Sc = Sa_1b$. By Lemma 2.1, $Pc = Pa_1b$, and in particular $c \in Pb$. Similarly, from $c^{-1}c = b^{-1}b \cdot a^{-1}a$ we infer that $c \in Pa$. Thus $Pc \subseteq Pa \cap Pb$. On the other hand, let $d \in Pa \cap Pb$. Then $d \in Sa \cap Sb = Se_a \cap Se_b$. Hence $de_a = de_b = d$, and therefore $de_c = de_a e_b = d$. Thus $d \in Se_c = Sc$, and, again using Lemma 2.1, $d \in Pc$. Hence $Pa \cap Pb \subseteq Pc$, and equality follows.

Assume conversely that $Pa \cap Pb = Pc$. As noted just before the statement of the lemma, there is an element x of P such that $e_a e_b = e_b e_a = e_x$. By the foregoing, $Px = Pa \cap Pb$. Hence $Px = Pc$, and $e_x = e_c$ by Lemma 3.1.

We collect these results into the following theorem, the proof of which is evident from Lemmas 3.1 and 3.2.

THEOREM 3.1. *Let S be a semigroup satisfying* A1, 2, 3, *and let P be its right unit subsemigroup. Then P satisfies* B3 *(as well as* B1 *and* B2*), and the semi-lattice of principal left ideals of P under intersection is isomorphic with the semi-lattice of idempotent elements of S.*

4. Proof of the main theorem. Let P be a semigroup satisfying conditions B1, 2, 3. Let U be the group of units of P. If a and b are elements of P, we shall write $a \sim b$ if a and b are left associate: $Pa = Pb$.

LEMMA 4.1. *If $a \sim b$ then $a = ub$ with $u \in U$.*

Proof. From $Pa = Pb$ we have $a = xb$ and $b = ya$ with $x, y \in P$. From $a = xya$, $b = yxb$, and B1, we conclude $xy = yx = 1$. Thus, $x, y \in U$.

We note incidentally that every left or right unit of P is a (two-sided) unit. For if $ab = 1$ then $(ba)^2 = ba$ and hence $ba = 1$, since by B1 the only idempotent element of P is 1.

We now suppose that a representative is chosen in each class of left associates of P, so that the LCLM $a \vee b$ becomes definite, as in § 1. The operation $*$ is defined by (1.1).

LEMMA 4.2. $(a \vee b)c \sim ac \vee bc \,(a, b, c \in P)$.

Proof. $P(a \vee b)c = (Pa \cap Pb)c = Pac \cap Pbc = P(ac \vee bc)$.

LEMMA 4.3. *For any element a of P, $u_a = 1 * a$ is a unit, and $a * 1 = a \vee 1 = u_a a$.*

Proof. $a \vee 1 \sim a$, and hence $a \vee 1 = u_a a$ with $u_a \varepsilon U$ by Lemma 4.1. From $(1 * a)a = 1 \vee a = a \vee 1 = u_a a$ and B1, we conclude $1 * a = u_a$. For the rest, $a * 1 = (a * 1)1 = a \vee 1$.

LEMMA 4.4. *For $a, b \varepsilon P$ and $u, v \varepsilon U$, we have $(ua * vb)v = a * b$.*

Proof. $ua \vee vb = a \vee b$ since both are equal to the representative element of the same class of associates. Hence

$$(ua * vb)vb = ua \vee vb = a \vee b = (a * b)b$$

and the result follows from B1.

LEMMA 4.5. *If $a' = ua$, $b' = ub$, $c' = vc$, $d' = vd$ with $u, v \varepsilon U$, then $(c' * b')a' = (c * b)a$ and $(b' * c')d' = (b * c)d$.*

Proof. By Lemma 4.4,

$$(c' * b')a' = (vc * ub)ua = (c * b)a,$$

$$(b' * c')d' = (ub * vc)vd = (b * c)d.$$

We now let $P^{-1} \circ P$ be the set of ordered pairs (a, b) of elements of P, with equality defined by (1.2), and define a product in $P^{-1} \circ P$ by (1.3). By Lemma 4.5, this product is single-valued.

LEMMA 4.6.

(1) $\quad (a, b)(1, c) = (a, bc),$ $\qquad$ (2) $\quad (a, 1)(1, c) = (a, c),$

(3) $\quad (1, b)(1, c) = (1, bc),$ $\qquad$ (4) $\quad (a, 1)(c, 1) = (ca, 1).$

Proof. By (1.2), (1.3), and Lemma 4.3,

$$(a, b)(1, c) = ((1 * b)a, (b * 1)c) = (u_b a, u_b bc) = (a, bc).$$

(2) and (3) follow from (1) on setting $b = 1$ and $a = 1$, resp. To show (4):

$$(a, 1)(c, 1) = ((c * 1)a, (1 * c)1) = (u_c ca, u_c) = (ca, 1).$$

LEMMA 4.7.

$$(a', a)(1, b) \cdot (c, c') = (a', a) \cdot (1, b)(c, c').$$

Proof. Let $p = c * b$, $q = c * ab$, $r = p * a$ and $p' = b * c$, $q' = ab * c$, $r' = a * p$, so that $pb = b \vee c = p'c$, $qab = ab \vee c = q'c$, $ra = a \vee p = r'p$. Then, by Lemma 4.2,

$$rab = (a \vee p)b \sim ab \vee pb \sim ab \vee b \vee c \sim ab \vee c = qab.$$

By Lemma 4. 1, $rab = uqab$ with $u \varepsilon U$, and hence $r = uq$ by B1. Furthermore;

$$r'p'c = r'pb = rab = uqab = uq'c, \text{ whence } r'p' = uq',$$

by B1. By Lemma 4. 6 (1),

$$(a', a)(1, b) \cdot (c, c') = (a', ab)(c, c') = ((c * ab)a', (ab * c)c') = (qa', q'c').$$

On the other hand,

$$\begin{aligned}(a', a) \cdot (1, b)(c, c') &= (a', a)(c * b, (b * c)c') = (a', a)(p, p'c') \\ &= ((p * a)a', (a * p)p'c') = (ra', r'p'c') = (uqa', uq'c') = (qa', q'c').\end{aligned}$$

LEMMA 4. 8. *The associative law holds in $P^{-1} \circ P$.*

Proof. By Lemma 4. 7, the associative law holds for any triad the middle one of which has the form $(1, b)$. The same is proved analogously if the middle term has the form $(b, 1)$. Using Lemma 4. 6 (2),

$$\begin{aligned}(a, a') \cdot (b, b')(c, c') &= (a, a')[(b, 1)(1, b') \cdot (c, c')] \\ &= \cdots = [(a, a') \cdot (b, 1)(1, b')](c, c') = (a, a')(b, b') \cdot (c, c'),\end{aligned}$$

where the dots indicate four applications of Lemma 4. 7 or its dual.

By Lemma 4. 6 (3), the elements $(1, a)$ of $P^{-1} \circ P$ constitute a subsemigroup thereof, isomorphic with P. The next lemma shows that this is just the right unit subsemigroup of $P^{-1} \circ P$.

LEMMA 4. 9. *The elements of $P^{-1} \circ P$ having right inverses are just those of the form $(1, a)$. The right inverse of $(1, a)$ is $(a, 1)$.*

Proof. By Lemma 4. 1, $a \vee a = ua$ with $u \varepsilon U$. From $(a * a)a = a \vee a = ua$, we conclude $a * a = u$. Hence $(1, a)(a, 1) = (a * a, a * a) = (u, u) = (1, 1)$. Suppose conversely that $(a, b)(c, d) = (1, 1)$, that is, $((c * b)a, (b * c)d) = (1, 1)$. This implies that $(c * b)a$, and hence also a, is a unit. But then $(a, b) = (1, a^{-1}b)$.

LEMMA 4. 10. *$P^{-1} \circ P$ is d-simple.*

Proof. By Lemma 4. 9 and its dual, together with Lemma 4. 6 (2), every element of $P^{-1} \circ P$ is expressible as the product of a left unit by a right unit. By Theorem 2. 1, $P^{-1} \circ P$ is d-simple.

LEMMA 4. 11. *The idempotent elements of $P^{-1} \circ P$ are just those of the form (a, a). We have $(a, a)(b, b) = (a \vee b, a \vee b)$ and, in particular,* A3 *holds in $P^{-1} \circ P$.*

Proof. Suppose (a, b) is idempotent:

$$(a, b) = (a, b)(a, b) = ((a * b)a, (b * a)b).$$

This means that $(a * b)a = ua$, $(b * a)b = ub$ for some $u \varepsilon U$. Hence $a * b = b * a = u$. From $(a * b)b = a \vee b = (b * a)a$ we conclude that $ub = ua$, or $b = a$.

Conversely, since $a * a \varepsilon U$, we have

$$(a, a)(a, a) = ((a * a)a, (a * a)a) = (a, a).$$

If now (a, a) and (b, b) are any two idempotents of $P^{-1} \circ P$, then

$$(a, a)(b, b) = ((b * a)a, (a * b)b) = (a \vee b, a \vee b).$$

This last equation, incidentally, shows again the isomorphism between the semi-lattice of all idempotent elements of $P^{-1} \circ P$ and that of P under LCLM.

The first part of the main theorem is established by Lemmas 4. 8-4. 11. Turning to the second half, Theorem 3. 1 shows that P satisfies B1, 2, 3. We proceed to show that S is isomorphic with $P^{-1} \circ P$. By (1. 2) and Lemma 2. 3, we see that the correspondence $a^{-1}b \leftrightarrow (a, b)$ is one-to-one between S and $P^{-1} \circ P$. To prove the isomorphism, we must show that two elements $a^{-1}b$, $c^{-1}d$ of S multiply like (1. 3), i. e., $a^{-1}b \cdot c^{-1}d = [(c * b)a]^{-1}[(b * c)d]$, where $*$ is defined by (1. 1). From $c * b = (c \vee b)b^{-1}$ we have $(c * b)^{-1} = b(c \vee b)^{-1}$ $= b(b \vee c)^{-1}$. By Lemma 3. 2, $e_{b \vee c} = e_b e_c$, and hence

$$\begin{aligned}[(c * b)a]^{-1}[(b * c)d] &= a^{-1}(c * b)^{-1} \ (b * c)d \\ &= a^{-1}b(b \vee c)^{-1} \cdot (b \vee c)c^{-1}d = a^{-1}b \cdot e_{b \vee c} \cdot c^{-1}d \\ &= a^{-1}b \cdot e_b e_c \cdot c^{-1}d = a^{-1}b \cdot c^{-1}d,\end{aligned}$$

since $be_b = b$ and $e_c c^{-1} = c^{-1}$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] Garrett Birkhoff, *Lattice Theory* (American Mathematical Society Colloquium Publications, vol. 25, New York, 1948).

[2] J. A. Green, "On the structure of semigroups," *Annals of Mathematics*, vol. 54 (1951), pp. 163-172.

[3] D. Rees, "On semi-groups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400.

[4] A. Suschkewitsch, "Über die endlichen Gruppen ohne das Gesetz der eindeutigen Umkehrbarkeit," *Mathematische Annalen*, vol. 99 (1928), pp. 30-50.

STRONGLY TOPOLOGICAL IMBEDDING OF F_σ-SUBSETS OF E_n.[1]

By Henry Sharp, Jr.

1. **Introduction.**[2] All spaces considered in this paper shall be separable and metric. Denote by E_n Euclidean space of n dimensions, by M_n^k the set of points in E_n having at most k rational coordinates, by L_n^k the set of points in E_n having at least k rational coordinates. Note that $L_n^k = E_n - M_n^{k-1}$, $\dim M_n^k = k$, $\dim L_n^k = n - k$ [5], p. 42. It is well known that the set M_{2k+1}^k contains a topological image of every k-dimensional point set and that the number $2k+1$ cannot be improved. It is, however, an open question whether M_n^k $(n < 2k+1)$ contains a topological image of every k-dimensional subset of E_n [5], p. 65. For the special cases $k = n$ and $k = n-1$ the answer to this question is in the affirmative; in fact, for these values of k the set M_n^k contains a *strongly topological image* (image under a homeomorphism of the space onto itself) of every k-dimensional subset of E_n.

The present paper is concerned with a related problem: does the set M_n^k contain a strongly topological image of every j-dimensional F_σ-subset of E_n? (Obviously the problem is solved for $k = n$ and $k = n - 1$.) S. W. Hahn has given the answer to this problem for certain special cases, namely: affirmative for $n = 2$, $k = 0$, and $n = 3$, $k = 1$, $j = 0$; negative for $n = 3$, $k = 0$ [4], pp. 308-311. The solution of this problem is completed in the present paper in which it is shown that the answer is affirmative for $n \geqq 4$, $k = n - 2$, and negative in all other previously unsolved cases. The results here together with previous results are concisely summarized in Figures 1 and 2. The presence of lines connecting a particular S_n^j with a particular M_n^k indicates that every j-dimensional F_σ-subset of E_n is strongly homeomorphic to a subset of M_n^k, while the absence of connecting lines indicates that there exists a compact j-dimensional subset of E_n which is not strongly homeomorphic to a subset of M_n^k. For example, the figure indicates that M_4^2 contains a strongly homeomorphic image of every 0-, 1-, and 2-dimensional

[1] Received September 9, 1952.

[2] This paper is taken from the author's Ph. D. dissertation, Duke University, 1952.

F_σ-subset of E_4, and also that there exist both 0-, and 1-dimensional compact sets in E_4 which are not strongly homeomorphic to a subset of $M_4{}^1$.

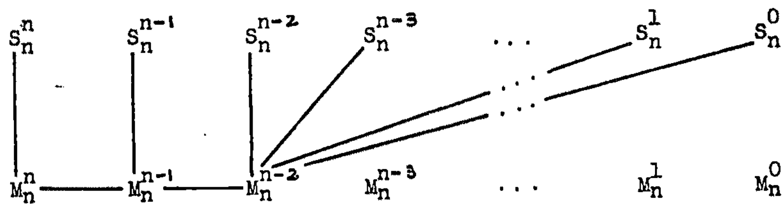

FIGURE 1 ($n = 0, n = 1, n = 2, n \geqq 4$).

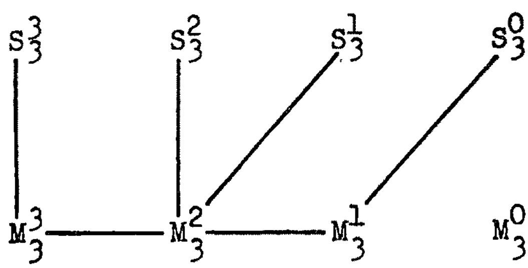

FIGURE 2.

The principal affirmative result is embodied in the theorem of Section **3**, while the negative results follow from the counter-examples discussed in Sections **4** and **5**. Section **2** is used to establish two lemmas which are essential in the proof of the theorem. Section **6** concludes the paper with an indication of several of the unsolved problems suggested by this investigation. The author wishes to acknowledge his indebtedness to Professor John H. Roberts for his guidance in the preparation of this paper.

2. Preliminary lemmas.

LEMMA 1. *Suppose A, B, C are distinct non-collinear points in E_n. Let $\epsilon > 0$ be given. Let $\alpha = [A, B] \cup [B, C]$, let $\lambda = [A, C]$, let σ be the 2-cell (A, B, C), let Δ be the set of points of E_n whose distance from σ is less than ϵ. Then there exists an n-polytope K_n and a homeomorphism $f(E_n) = E_n$ such that (1) $\sigma \subset K_n \subset \Delta$, (2) $f(\alpha) = \lambda$, (3) $f(P) = P$ for P belonging to $E_n - K_n$.*

Proof. We give a proof of this lemma for the case $n = 3$, and it is

evident that a continuation of the method of proof will establish the lemma for any integer $n > 3$.

In E_3 let the system of coordinates be chosen so that λ is on the x_1 axis with its midpoint, Q, at the origin of coordinates, and so that B is in the x_1x_2-plane with $B^2 > 0$ (P^i denotes the i-th coordinate of the point P). Let l be the line on the points B and Q, and let D ($D^2 > B^2$) and E ($E^2 < 0$) be points of l contained in Δ. Let K_2 denote the 2-polytope whose vertices are A, D, B, Q, E, C (as in Figure 3).

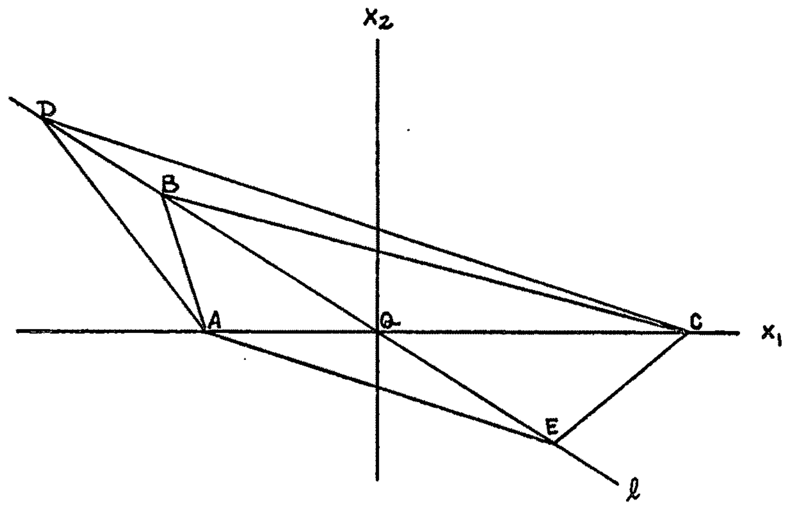

FIGURE 3.

There exists a piecewise linear transformation f_1 of the plane x_1x_2 onto itself which moves only points interior to K_2 such that $f_1[A, B] = [A, Q]$ and $f_1[B, C] = [Q, C]$, and such that $f_1(P)$ is on the line through P parallel to l.

Let F ($F^3 > 0$) and G ($G^3 < 0$) be points on the x_3-axis lying in Δ, and let K_3 be the 3-polytope which is the smallest convex set containing K_2, F, and G. There exists a piecewise linear transformation f_2 of E_3 onto itself which moves only points interior to K_3 such that $f_2(P) = f_1(P)$ if P belongs to the plane x_1x_2, and such that $f_2(P)$ is on the plane through P parallel to the x_1x_2-plane. The transformation thus described is a homeomorphism satisfying the lemma for the case $n = 3$.

If the points A, B, C are related in such a way that $A^1 < B^1 < C^1$, then by a slight modification in the above proof it is possible to add a fourth

conclusion to the lemma: (4) *if P lies on an $(n-1)$-dimensional hyperplane perpendicular to λ, then $f(P)$ lies on the same hyperplane.*

LEMMA 2. *Let X be a compact point set of dimension $\leqq n-2$ contained in E_n, $n \geqq 4$. Let L be a line in E_n and let $\epsilon > 0$ be assigned. Then there exists a homeomorphism $h(E_n) = E_n$ and a compact set D such that (1) $h(X) \cap L = 0$, (2) $d(P, h(P)) < \epsilon$ for all points P of E_n, (3) $h(P) = P$ for all points P of $E_n - D$.*

Proof. For convenience assume that L is the x_1-axis. There exists a positive number η such that $X \cap L$ is contained in the segment $(-\eta, \eta)$ on L. Let m be a positive integer so large that $2\eta/m < \epsilon/n^{\frac{1}{2}}$, and let the interval $[-\eta, \eta]$ be divided into m equal subintervals each of length δ $(= 2\eta/m)$ by the $m+1$ points $A_0, A_1, \cdots, A_m$ ordered from $A_0\colon(-\eta, 0, \cdots, 0)$ to $A_m\colon(\eta, 0, \cdots, 0)$. Let D_i, $i = 1, 2, \cdots, m$, represent the n-dimensional cube bounded by the $(n-1)$-dimensional hyperplanes $x_1 = a_{i-1}$, $x_1 = a_i$, $x_2 = \pm \frac{1}{2}\delta$, $x_3 = \pm \frac{1}{2}\delta, \cdots, x_n = \pm \frac{1}{2}\delta$, where a_i is the x_1 coordinate of the point A_i. Let $D = \bigcup_{i=1}^{m} D_i$ and let P_i, $i = 1, 2, \cdots, m-1$, be a point contained in the interior of $D - X$ lying on the hyperplane $x_1 = a_i$. For convenience of notation let $P_0 = A_0$, $P_m = A_m$. For each i, $i = 1, 2, \cdots, m$, we define a homeomorphism f_i of E_n onto itself which is the identity outside of D_i. The number m was so chosen that the diameter of D_i is less than ϵ. Since X cannot separate D_i, there exists a simple polygonal arc α_i from P_{i-1} to P_i such that (1) $X \cap \alpha_i = 0$, (2) except for end points α_i is contained in the interior of D_i, (3) the vertices of α_i are in general position; that is, no $j+2$ of the vertices lie in a j-dimensional linear subspace of E_n, $j = 0, 1, \cdots, n-1$. By using Lemma 1 a finite number of times on the vertices of α_i, each use of Lemma 1 yielding an arc having one less vertex than the previous arc, we determine a homeomorphism f_i of E_n onto itself which brings α_i into coincidence with the line interval $[P_{i-1}, P_i]$ and which moves only points interior to D_i. It should be noted here that the condition $n \geqq 4$ is essential to insure that no use of Lemma 1 shall disturb points of the arc other than those specifically under consideration. If $f = \prod_{i=1}^{m} f_i$ and $\beta = \bigcup_{i=1}^{m} [P_{i-1}, P_i]$ then (1) $f(\bigcup_{i=1}^{m} \alpha_i) = \beta$, (2) f is the identity on $E_n - D$, (3) if P is a point of D_i then $f(P)$ is a point of D_i, (4) $\beta \cap f(X) = 0$. Again there exists a homeomorphism g of E_n onto itself which is the resultant of a finite number of homeomorphisms, each leaving the x_1 coordinate invariant (applying here

conclusion 4 of Lemma 1), having the properties (1) $g(\beta) = D \cap L$, (2) g is the identity on $E_n - D$. Hence $h = gf$ is a homeomorphism of E_n onto itself such that (1) $L \cap h(X) = 0$, (2) $d(P, h(P)) < \epsilon$, (3) $h(P) = P$ for all points P of $E_n - D$.

3. The case $n \geqq 4$, $k = n - 2$.

THEOREM.[3] *The set M_n^{n-2}, $n \geqq 4$, contains an image under a homeomorphism $h(E_n) = E_n$ of every F_σ-subset of E_n of dimension $\leqq n-2$. Furthermore, the collection $\{h\}$ of all such homeomorphisms for a given F_σ forms a dense G_δ-subset of the set of all homeomorphisms of E_n onto itself.*

The proof of this theorem rests on the fact that L_n^{n-1} $(= E_n - M_n^{n-2})$ consists of a countable number of lines which can be imbedded under a strong homeomorphism in the complement of any F_σ-subset of E_n of dimension $\leqq n-2$. Before proceding to the proof, a few preliminary remarks are necessary. Let F_n denote the set of all mappings of the n-sphere, S_n, into itself which leave a given point, ω, invariant. We metrize F_n by the distance formula $\rho(f_1, f_2) = \max_{x \varepsilon S_n} \delta(f_1(x), f_2(x))$, where f_1, f_2 are elements of F_n and δ is the usual metric on the sphere. Since F_n is a closed subset of the complete space $S_n^{S_n}$ it follows that F_n is complete. Let H_n denote the set of all homeomorphisms of S_n onto itself leaving ω invariant. By an argument similar to one appearing in [5], p. 57 it follows that H_n is a G_δ-subset of F_n. Hence H_n can be assigned a metric, σ, under which it becomes a complete space (see, e. g., [7], p. 29). It is well known that there is a natural one-to-one correspondence, defined by the stereographic projection, between the elements of H_n and the homeomorphisms of E_n onto itself. Hereafter we will think of H_n as being the space of all homeomorphisms of E_n onto itself as metrized by σ or by the topologically equivalent metric ρ. (It should be noted that ρ, in contradistinction to σ, is not a complete metric). We note here for future reference that if x, y are points of E_n and x', y' the corresponding points of S_n then $\delta(x', y') \leqq 2 \cdot d(x, y)$.

LEMMA. *Let X be a compact subset of E_n, $n \geqq 4$, of dimension $\leqq n-2$, and let L be a line in E_n. Then the set G of all homeomorphisms $g(E_n) = E_n$ such that $g(X) \cap L = 0$ is both open and dense in H_n.*

Proof. If g belongs to G then since $g(X)$ is compact and L is a line

---

[3] This theorem, as originally stated, included only the first sentence. The author is indebted to Professor S. Eilenberg and to the referee for their indication that Baire's Theorem could be applied to yield the stronger result.

it follows by an elementary argument that G is open in H_n. Let h be an arbitrary element of H_n and let $\epsilon > 0$ be assigned. We wish to show that there exists an element g of G such that $\sigma(h, g) < \epsilon$. But since ρ and σ are topologically equivalent metrics and density is a topological property, it follows that it will be sufficient to determine g belonging to G such that $\rho(h, g) < \epsilon$. Since $h(X)$ is compact there exists, by Lemma 2 of Section 2, an element ϕ of H_n such that $\phi[h(X)] \cap L = 0$ and $d(x, \phi(x)) < \frac{1}{2}\epsilon$ for all points x, where d is the Euclidean metric. Since $\delta(x', \phi'(x)) < 2 \cdot d(x, \phi(x))$ it follows that $\rho(e, \phi) < \epsilon$, e denoting the identity element of H_n. But $\rho(h, \phi h) = \rho(e, \phi)$, hence $\rho(h, \phi h) < \epsilon$. To complete the proof we have only to set $g = \phi h$.

Proof of the theorem. Let X be an F_σ-subset of E_n of dimension $\leqq n-2$. Put $X = \bigcup_{i=1}^{\infty} X_i$, where for each i the set X_i is compact and of dimension $\leqq n-2$. The set L_n^{n-1} is the union of a countable number of lines, $L_n^{n-1} = \bigcup_{j=1}^{\infty} L_j$. Let $H^{i,j}$ denote the set of all elements h of H_n such that $h(X_i) \cap L_j = 0$. By the previous lemma $H^{i,j}$ is both open and dense in H_n. Hence, applying Baire's Theorem, the intersection of all the $H^{i,j}$ is a dense G_δ-set in H_n.

4. The case $n \geqq 4$, $k < n-2$.

LEMMA. *For $n \geqq 2$ the identity mapping f of the space L_n^{n-1} in the space L_n^{n-2} is inessential.*

Indication of proof. We wish to show that f is homotopic to zero in L_n^{n-2}. For p: $(x_1, x_2, \cdots, x_n)$ belonging to L_n^{n-1}, $0 \leqq t \leqq 1$, we define a homotopy θ_1 as follows: $\theta_1(p, t) = ((1-t)x_1, x_2, \cdots, x_n)$. Since p has at least $n-1$ rational coordinates it follows that $\theta_1(p, t)$ has at least $n-2$ rational coordinates, hence is in L_n^{n-2}. Clearly $\theta_1(p, 0) = f(p) = p$, $\theta_1(p, 1) = (0, x_2, \cdots, x_n)$ which is a point of L_n^{n-1}. Similarly, we can define a homotopy θ_2 over the set $\theta_1(L_n^{n-1}, 1)$, $0 \leqq t \leqq 1$, so that $\theta_2[\theta_1(p, 1), 1] = (0, 0, x_3, \cdots, x_n)$. A series of n such homotopies can be defined leading to the desired result.

Antoine has given an example of a compact, 0-dimensional subset of E_3 whose complement is not simply connected [1], 663. Recently, W. A. Blankinship, through a generalization of Antoine's example, has shown the existence of a compact, 0-dimensional subset of E_n, $n \geqq 3$, whose complement

is not simply connected [2]. Blankinship's result leads to the following theorem: *Not every compact, 0-dimensional subset of E_n, $n \geqq 4$, is strongly homeomorphic to a subset of M_n^{n-3}.* For let A be a compact, 0-dimensional subset of E_n such that $E_n - A$ is not simply connected. If there exists a strong homeomorphism h such that $h(A)$ is contained in M_n^{n-3}, then $E_n - h(A)$ contains L_n^{n-2}. The property of being simply connected is a topological invariant, hence there exists a closed path J in $E_n - h(A)$ which is not homotopic to zero in $E_n - h(A)$. But L_n^{n-2} consists of a countable number of planes (parallel to the coordinate planes) which are dense throughout the space. It is evident that J is homotopic in $E_n - h(A)$ to a closed path J' contained in L_n^{n-1}, and since the identity mapping of L_n^{n-1} in L_n^{n-2} is inessential, the path J must be homotopic to zero.

5. The case $n = 3$, $k = 1$. Previous results have completely solved the problem under consideration for the cases $n = 0$, $n = 1$, and $n = 2$, and have left just one question for the case $n = 3$, namely: does the set M_3^1 contain a strongly topological image of every 1-dimensional F_σ-subset of E_3? That the answer to this question is in the negative is shown by an example of a compact 1-dimensional subset of E_3 described in a note added by the referee to a paper of Frankl and Pontrjagin [3], p. 788. This example is obtained by omitting from the unit cube a countably infinite sequence of knotted canals running between opposite faces of the cube. Denote this compact, 1-dimensional set by D and suppose D can be imbedded in M_3^1 by a strong homeomorphism. Then by taking complements L_3^2 is contained in $E - f(D)$. But this is impossible since L_3^2 is the set of rational lines contained in E_3.

6. Conclusion. The only case which differs from the general pattern of results in Figure 1 is the negative result for 1-dimensional compact subsets of E_3. This divergence is intimately connected with the existence of knotted curves, which can occur only in a space of three dimensions. Thus the proof of the lemma of Section 2 breaks down if $n = 3$ since then there is no guarantee that the closed curve $\alpha_i + [P_{i-1}, P_i]$ will be free of knots.

This investigation has left unsolved several problems, of a similarly specialized nature, in connection with the general problem stated in the first paragraph of the introduction. It would be desirable to find an answer to the question: does there exist a k-dimensional subset of E_n which contains a strongly homeomorphic image of every k-dimensional compact subset of E_n, $n \geqq 3$, $k < n - 2$? Sections 4 and 5 prove that the set M_n^k does not have

this property. Another question arises on removing the restriction of strong homeomorphisms: does the set $M_n{}^k$, $n \geqq 6$, $[\frac{1}{2}n - \frac{1}{2}] < k < n - 2$ contain a topological image of every k-dimensional F_σ-subset of E_n?

DUKE UNIVERSITY.

BIBLIOGRAPHY.

[1] L. Antoine, "Sur la possibilité d'étendre l'homéomorphie de deux figures à leur voisinage," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 171 (1920), pp. 661-663.

[2] W. A. Blankinship, "Generalization of a construction of Antoine," *Annals of Mathematics* (2), vol. 53 (1951), pp. 276-297.

[3] F. Frankl and L. Pontrjagin, "Ein Knotensatz mit Anwendung auf die Dimensiontheorie," *Mathematische Annalen*, vol. 102 (1930), pp. 785-789.

[4] S. W. Hahn, "Universal spaces under strong homeomorphisms," *Transactions of the American Mathematical Society*, vol. 70 (1951), pp. 301-311.

[5] W. Hurewicz and H. Wallman, "*Dimension Theory*," revised edition, Princeton, 1948.

[6] J. H. Roberts, "Concerning topological transformations in E_n," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 252-262.

[7] G. T. Whyburn, "Analytic Topology," *American Mathematical Society Colloquium Publication*, vol. 28, 1942.

ON THE CHARACTERISTIC CLASSES OF COMPLEX SPHERE BUNDLES AND ALGEBRAIC VARIETIES.* †

By Shiing-shen Chern.

Introduction. In a recent paper [1] Hodge studied the question of identifying, for non-singular algebraic varieties over the complex field, the characteristic classes of complex manifolds [2] with the canonical systems introduced by M. Eger and J. A. Todd.[3] He proved that they are identical up to a sign, when the algebraic variety is the complete intersection of non-singular hypersurfaces in a projective space. His method does not seem to extend to a general algebraic variety. One of the main difficulties lies in the fact that the theory of canonical systems of algebraic varieties has so far been developed only in broad outlines, with the result that very few of their properties are available.

We shall give in this paper a more direct treatment of the problem, by proving that there is an equivalent definition of the characteristic classes, which is valid for algebraic varieties. In order to make the paper as self-contained as possible, let us begin by recalling the original definition of the characteristic classes. We consider a compact complex manifold M_n [4] of complex dimension n, and over M_n consider the bundle B_{nr}^* of ordered sets $(e_1, \cdots, e_r)$ of r linearly independent complex vectors with the same origin.[5] The fiber of this bundle is the complex Stiefel manifold V_{nr} of all the ordered sets of r linearly independent complex vectors in a complex vector space of dimension n. It is well-known that V_{nr} is connected and that its first non-vanishing homotopy group is $\pi_{2n-2r+1}(V_{nr})$, the latter being free cyclic. To describe a generator of $\pi_{2n-2r+1}(V_{nr})$ we fix $e_1, \cdots, e_{r-1}$ and let W_{n-r+1}

* Received September 29, 1952.

† This work is done under partial support of the Office of Naval Research.

[1] Hodge [9]. The number refers to the bibliography at the end of the paper.

[2] Chern [1].

[3] Eger [2] and Todd [13].

[4] In general, we shall use a subscript to denote complex dimension and a superscript to denote topological dimension. When the meaning is clear, it will be dropped to simplify notation.

[5] While we shall explain, in so far as possible, the notions which will be utilized, we shall refer to Steenrod [12] as our standard reference on fiber bundles.

be a vector space of dimension $n-r+1$ which contains no non-trivial linear combination of $e_1, \cdots, e_{r-1}$. Let W_{n-r+1}^* be the space obtained from W_{n-r+1} by deleting its origin. Then $\pi_{2n-2r+1}(V_{nr})$ and $\pi_{2n-2r+1}(W_{n-r+1}^*)$ are naturally isomorphic (under the homomorphism induced by the inclusion mapping $W_{n-r+1}^* \subset V_{nr}$). Since W_{n-r+1}, as a complex vector space, is oriented, its orientation determines uniquely a generator of $\pi_{2n-2r+1}(W_{n-r+1}^*)$, and hence of $\pi_{2n-2r+1}(V_{nr})$. In other words, the bundle B_{nr}^* is orientable. It follows from the theory of obstructions that the primary obstruction of this bundle is a cohomology class C_{n-r+1} of (topological) dimension $2(n-r+1)$ with integer coefficients. We call C_r, $r=1, \cdots, n$, the *r-th characteristic class* of M. If M also denotes the fundamental homology class of the oriented manifold M, the homology class γ_r defined by the cap product $\gamma_r = C_{n-r} \cap M$ is called the *r-th characteristic homology class* of M.

All these considerations apply to the case in which M is a non-singular algebraic variety over the complex field. However, since the obstructions are defined in terms of continuous cross sections over the skeletons of a triangulation of M, it does not follow that the characteristic homology class γ_r contains as representative an algebraic cycle, that is, a cycle in the form of a finite sum $\Sigma\lambda_i V_r^i - \Sigma\mu_k V_r^k$, where V_r^i, V_r^k are algebraic sub-varieties of dimension r in M and $\lambda_i \geqq 0$, $\mu_k \geqq 0$. A main purpose of this paper is to prove that this is the case.

We proceed to enumerate our results, postponing their proofs for later sections. In the course of our discussion several theorems on the homology theory of fiber bundles and on complex sphere bundles will be used. While they are to some extent known, they are either not easily accessible or not given in a form needed for our purpose. For the sake of completeness such results will be included here.

We consider a fiber bundle $p: B \rightarrow X$, about which we make once for all the following assumptions: 1) the base space X is a finite polyhedron; 2) the fiber F is a connected finite polyhedron; 3) the fundamental group of X acts trivially on the homology groups of the fibers under consideration. The last assumption makes it possible to use these groups as coefficient groups in the homology of X. Let $r > 1$ be such that $\pi_r(F) \neq 0$, $\pi_s(F) = 0$ for all $s < r$. Then it is well-known that $\pi_r(F)$ is isomorphic to the homology group $H_r(F)$ with integer coefficients. The primary obstruction of the bundle is an element of $H^{r+1}(X, H_r(F))$. Its vanishing has an implication described by the following theorem:

THEOREM 1. *Let $\pi_r(F) \cong H_r(F)$, $r > 1$, be the first non-vanishing*

homotopy group of F. If the primary obstruction vanishes, the injection mapping $l: F \to B$ induces a homomorphism $l^*: H^r(B, H_r(F)) \to H^r(F, H_r(F))$, which is onto.

The conclusion of this theorem gives information on the "homology position" of a fiber in the bundle. Relative to a coefficient group G the simplest situation is when F is totally non-homologous to zero, i. e., when the homomorphism $l^*: H^r(B, G) \to H^r(F, G)$ is onto for all r. When G is a field, such bundles were studied by Leray and Hirsch, and the cohomology ring of the bundle is found to be isomorphic, in its additive structure, to that of the Cartesian product of the fiber and the base space. Since we are mainly interested in integer coefficients, we need the following strengthened form, due to E. H. Spanier, of the Leray-Hirsch theorem:[6]

THEOREM 2. *Let $l: F \to B$ be the injection of the fiber into the bundle. Relative to a simple coefficient system G suppose there is a homomorphism $\mu: H^r(F; G) \to H^r(B; G)$ such that $l^*\mu$ is the identity automorphism of $H^r(F; G)$ for all $r \geqq 0$. Then $H^r(B; G)$ is isomorphic with $H^r(X \times F; G)$.*

We now describe an important operation in the homology theory of fiber bundles, known as "integration over the fiber." Let $H^r(F)$, $r > 0$, be the last non-vanishing cohomology group of the fiber, so that $H^s(F) = 0$ for all $s > r$. If G is a simple system of coefficient groups (that is, a system of local groups in B on which $\pi_1(B)$ acts trivially), integration over the fiber is a homomorphism

$$\Psi: H^m(B; G) \to H^{m-r}(X; H^r(F; G)). \tag{1}$$

To define Ψ let X be triangulated and let X^k be its k-dimensional skeleton. Put $B_k = p^{-1}(X^k)$. Then it is easy to see that

$$H^m(B_s; G) = 0, \qquad 0 \leqq s \leqq m - r - 1. \tag{2}$$

From the exact sequence of the pair (B_{m-r}, B_{m-r-1}):

$$\cdots \to H^m(B_{m-r}, B_{m-r-1}; G) \xrightarrow{j^*} H^m(B_{m-r}; G) \xrightarrow{i^*} H^m(B_{m-r-1}; G) \xrightarrow{\delta^*} \cdots, \tag{3}$$

it follows that j^* is onto. To an element $u \in H^m(B; G)$ let

$$u' = i^* u \in H^m(B_{m-r}; G)$$

be the image of u under the dual homomorphism of the homomorphism induced by $i: B_{m-r} \to B$. Since j^* is onto, there exists $v \in H^m(B_{m-r}, B_{m-r-1}; G)$

[6] Hirsch [7], Leray [10], pp. 183-184, and Spanier [11].

such that $j^*v = u'$. By an isomorphism which will later be described in more detail, we see that v can be identified with an $(m-r)$-dimensional cochain of X, with coefficients in $H^r(F;G)$. It can be proved that it is a cocycle and that its cohomology class depends only on u. This class is defined to be Ψu.

This operation Ψ has a simple geometrical interpretation, when X, B, F are oriented manifolds. In this case $H^r(F;G)$ is naturally isomorphic to G, so that the coefficient group on the right-hand side of (1) can be replaced by G. Denote also by X, B, F the fundamental homology classes of these manifolds. For $u \in H^m(B;G)$, $v \in H^s(X;G)$ we define, by means of cap products, the operations

$$\begin{aligned} \mathfrak{D}_B u &= u \cap B, \\ \mathfrak{D}_X v &= v \cap X. \end{aligned} \tag{4}$$

Then we have the theorem:

THEOREM 3. *If the spaces X, B, F of a fiber bundle are oriented manifolds and $u \in H^m(B;G)$, we have*

$$p_* \mathfrak{D}_B u = \mathfrak{D}_X \Psi u. \tag{5}$$

In other words, integration over the fiber is in this case dual to the homomorphism of homology classes induced by the projection p.

The results we need next center around the theory of complex sphere or vector bundles. As the structural group of the bundle of tangent vectors of M is the general linear group $G(n)$ in n complex variables, there is an associated bundle corresponding to every subgroup of $G(n)$. We realize $G(n)$ as the group of all $n \times n$ non-singular matrices with complex elements. Let $H(n,r)$ be the subgroup of $G(n)$, consisting of all matrices (a_{ik}) for which

$$a_{ik} = 0, \quad 1 \leqq k < i \leqq r; \quad r+1 \leqq i \leqq n, \quad 1 \leqq k \leqq r, \tag{6}$$

that is, of all matrices of the form

$$\begin{pmatrix} a_{11} \cdots a_{1r} & a_{1r+1} \cdots a_{1n} \\ \ddots & \\ 0 \quad a_{rr} & a_{r,r+1} \cdots a_{rn} \\ & a_{r+1,r+1} \cdots a_{r+1,n} \\ 0 & \cdots \\ & a_{n,r+1} \cdots a_{nn} \end{pmatrix}.$$

Let $K(n,r)$ be the subgroup of $H(n,r)$ whose matrices satisfy the further conditions $a_{11}=\cdots=a_{rr}=1$. Finally, let $L(n,r)$ be the subgroup of $K(n,r)$ whose matrices satisfy the additional conditions $a_{ik}=0$, $1\leqq i<k\leqq r$. We denote by B_{nr}, $\tilde{B}_{nr}$, and $B_{nr}{}^*$ the associated bundles corresponding to the subgroups $H(n,r)$, $K(n,r)$, and $L(n,r)$ respectively. The bundle $B_{nr}{}^*$ is the one introduced above, whose points are ordered sets of r linearly independent vectors $e_1(x),\cdots,e_r(x)$ with the same origin $x\in M$. Similarly, a point of $\tilde{B}_{nr}$ can be identified with a sequence of simple multivectors of the form $e_1(x), e_1(x)\wedge e_2(x),\cdots, e_1(x)\wedge\cdots\wedge e_r(x)\neq 0$, with the same origin $x\in M$. To describe the geometrical meaning of the bundle B_{nr} we need the notion of a tangent direction, which is the class of non-zero tangent vectors differing from each other by a non-zero complex factor. All the tangent directions at a point form a complex projective space of dimension $n-1$. A point of B_{nr} can be regarded as a sequence of linear spaces of directions $L_0(x)\subset L_1(x)\subset\cdots\subset L_{r-1}(x)$ in the space of tangent directions at $x\in M$, with the subscripts indicating the dimensions of these linear spaces. We notice that these spaces are related by natural projections as follows:

$$B_{nr}{}^* \xrightarrow{m_{nr}} \tilde{B}_{nr} \xrightarrow{p_{nr}} B_{nr} \xrightarrow{q_{nr}} M. \tag{7}$$

Moreover, under these projections every space is a bundle over the spaces which follow it.

It is the bundle

$$q_{nr}\colon B_{nr}\to M, \tag{8}$$

which we are most interested. Since $H(n,r)$ is defined for $1\leqq r\leqq n-1$, we shall assume r to be restricted by these inequalities, thus excluding $r=n$. The fiber F_{nr} of B_{nr} is the space of all sequences of linear subspaces $L_0\subset L_1\subset\cdots\subset L_{r-1}$ in a complex projective space of dimension $n-1$. Its cohomology ring with integer coefficients is generated by r two-dimensional cohomology classes and its first non-vanishing homology group of dimension >0 is $H_2(F_{nr})$, which is free abelian with r generators. Let M' be the $(2n-2r+1)$-dimensional skeleton of M. The bundle $B_{nr}{}^*$ has a cross-section $f\colon M'\to B_{nr}{}^*$ over M'. Then $p_{nr}\circ m_{nr}\circ f$ defines a cross-section of the bundle B_{nr} over M'. Since $2n-2r+1\geqq 3$, the primary obstruction of the bundle B_{nr} is zero. It follows from Theorem 1 that the homomorphism

$$l_{nr}{}^*\colon H^2(B_{nr};H_2(F_{nr}))\to H^2(F_{nr};H_2(F_{nr}))$$

induced by the inclusion mapping $l_{nr}\colon F_{nr}\to B_{nr}$ is onto. Since $H_2(F_{nr})$ is a free abelian group with r generators, the cohomology groups $H^2(F_{nr})$, $H^2(B_{nr})$

with integer coefficients can be imbedded isomorphically into $H^2(F_{nr}; H_2(F_{nr}))$, $H^2(B_{nr}; H_2(F_{nr}))$ respectively. It follows that the induced homomorphism $l_{nr}^*: H^2(B_{nr}) \to H^2(F_{nr})$ is also onto.

This induced homomorphism has another geometrical interpretation. In fact, the fiber of the bundle $p_{nr}: \tilde{B}_{nr} \to B_{nr}$ is a Cartesian product of r complex lines, each with the origin deleted. This bundle gives rise in the base space B_{nr} to r 2-dimensional characteristic classes with integer coefficients. It can be seen that their images under l_{nr}^* generate the cohomology ring of F_{nr}. Since l_{nr}^* is multiplicative, we see that the conditions of Theorem 2 are satisfied. This leads to the conclusion that the cohomology groups of B_{nr} are isomorphic to those of $M \times F_{nr}$. In other words, the space B_{nr} has rather simple additive homology properties.

In the study of the bundles in (7) an important tool is the so-called duality theorem. To describe the situation in geometrical terms, we take, over the same base space X, two bundles B_1, B_2 of complex vector spaces of dimensions ν_1, ν_2 respectively and construct a bundle of complex vector spaces of dimension $\nu_1 + \nu_2$ over X by taking as the fiber at a point $x \varepsilon X$ the space spanned by the fibers at x of the given bundles. This bundle will be called the product of B_1 and B_2 and will be denoted by $B_1 \boxtimes B_2$. The question naturally arises as to express the characteristic classes of $B_1 \boxtimes B_2$ in terms of those of B_1 and B_2. To express this relationship we introduce, for a bundle of complex vector spaces of dimension ν, the *characteristic polynomial*

$$C(t) = \sum_{t=0}^{\nu} C_i t^i, \qquad C_0 = 1. \tag{9}$$

This is a polynomial in an auxiliary variable t, whose coefficients are the characteristic cohomology classes C_i, with the convention that the classes of dimension greater than the topological dimension of X are replaced by zero. Then we have the theorem:

THEOREM 4. (Duality theorem for complex vector bundles) *If B_1 and B_2 are two complex vector bundles over the same base space X and if $C^{(1)}(t)$ and $C^{(2)}(t)$ are their characteristic polynomials, then the characteristic polynomial of their product bundle $B_1 \boxtimes B_2$ is $C^{(1)}(t)\ C^{(2)}(t)$.*

Using the duality theorem it is easy to prove the following theorem of G. Hirsch and Wu Wen-Tsun,[7] which can be regarded as giving a new definition of the characteristic classes:

[7] Hirsch [8]; Wu, unpublished.

THEOREM 5. *Let M be a compact complex manifold of dimension n, and B_{n1}, $\tilde{B}_{n1}$ two of the associate bundles of its tangent bundle, as defined above. Let u_1 be the characteristic class of the bundle $p_{n1}\colon \tilde{B}_{n1} \to B_{n1}$. Then we have*

$$(10) \qquad u_1{}^n = \sum_{i=1}^{n} (-1)^{i+1} q_{n1}{}^*(C_i) u_1{}^{n-i}.$$

This formula enables us to define the characteristic classes within the framework of homology theory. Unfortunately it does not seem to achieve our purpose of giving a definition applicable to the case when M is an algebraic variety. To obtain still another definition let us notice that the fiber F_{nr} of the bundle B_{nr} is an oriented manifold of topological dimension $r(2n-r-1)$, so that $H^{r(2n-r-1)}(F_{nr}; G)$ is naturally isomorphic to G and the homomorphism (1) can be written

$$(11) \qquad \Psi_{nr}\colon H^m(B_{nr}) \to H^{m-r(2n-r-1)}(M).$$

Moreover, instead of the characteristic polynomial $C(t)$ we can introduce the dual characteristic polynomial

$$(12) \qquad \bar{C}(t) = \sum_{i=1}^{n} \bar{C}_i t^i, \qquad \bar{C}_0 = 1,$$

defined by the condition

$$(13) \qquad C(t)\bar{C}(t) = 1.$$

Obviously the polynomial $C(t)$ defines $\bar{C}(t)$, and vice versa.

In order to formulate our next theorem, we need some notations. We regard a point of $\tilde{B}_{nr}$ as a sequence of simple multivectors of the form $e_1(x), \cdots, e_1(x) \wedge \cdots \wedge e_r(x) \neq 0$, with the same origin $x \in M$. Then the sequences of multivectors having the same projection in B_{nr} are exactly the ones obtained from the last sequence by multiplying its multivectors by the non-zero complex numbers $\alpha_1, \cdots, \alpha_r$ respectively. We can therefore regard $\alpha_1, \cdots, \alpha_r$ as the coordinates in the fiber of the bundle $p_{nr}\colon \tilde{B}_{nr} \to B_{nr}$. The fiber is thus a Cartesian product of r complex centered affine lines, each with the origin deleted. We denote by $v_1, \cdots, v_r$ their characteristic classes in B_{nr}. From them we introduce the cohomology classes $u_1, \cdots, u_r$ according to the equations

$$(14) \qquad v_i = u_1 + \cdots + u_i, \qquad i = 1, \cdots, r.$$

Then we have the theorem:

THEOREM 6. *For $1 \leqq r \leqq n-1$ the following formula holds:*

$$(15) \qquad \Psi_{nr}(u_1{}^{n-2} \cdots u_{r-1}{}^{n-r} u_r{}^{2n-r}) = (-1)^n \bar{C}_{n-r+1}.$$

In terms of homology this formula gives, on account of Theorem 3,

$$
\begin{aligned}
(-1)^n \mathcal{D}_M \bar{C}_{n-r+1} &= (p_{nr})_* \mathcal{D}_B (u_1^{n-2} \cdots u_{r-1}^{n-r} u_r^{2n-r}) \\
&= (p_{nr})_* \{ (\mathcal{D}_B u_1)^{n-2} \cdots (\mathcal{D}_B u_{r-1})^{n-r} (\mathcal{D}_B u_r)^{2n-r} \},
\end{aligned} \tag{16}
$$

where we write B for B_{nr}. Multiplication in the last expression means intersection of the homology classes.

Now let M be a non-singular algebraic variety in a complex projective space of higher dimension. It is well-known that B_{nr} is a non-singular algebraic variety and that the projection p_{nr} is a rational mapping. As first shown by Weil,[8] each of the classes $\mathcal{D} u_i$, $i = 1, \cdots, r$, in B_{nr} contains a divisor class which consists of all divisors linearly equivalent to each other. Since the intersection of divisor classes always contains an algebraic cycle and since, under a rational mapping, an algebraic cycle goes into an algebraic cycle, it follows from (16) that $\mathcal{D}_M \bar{C}_{n-r+1}$ contains an algebraic cycle, for $r = 1, \cdots, n-1$. On the other hand, $\mathcal{D}_M \bar{C}_1$ is a homology class which contains a divisor class of divisors. We can therefore state the theorem:

THEOREM 7. *Every characteristic homology class on a non-singular algebraic variety contains an algebraic cycle.*

1. On the homology theory of fiber bundles. The homology theory of fiber bundles has been the object of study of many authors. The problem is by nature not a simple one; published accounts of it are either sketchy or need much machinery in algebraic topology. We shall give below a procedure developed by E. H. Spanier and the author[9] which has the advantage of being quite elementary and which will lead to proofs of our Theorems 1, 2, and 3. We begin by discussing some elementary facts on the homology theory of topological spaces.

Let X be a topological space, and A, B, C, D four closed subsets, such that

$$D \subset C \subset A, \qquad D \subset B \subset A.$$

Then the inclusion mapping

$$j\colon\ (C, D) \subset (A, B)$$

induces a homomorphism of the relative cohomology groups:

$$j^*\colon\ H^r(A, B) \to H^r(C, D). \tag{17}$$

[8] Weil [16].

[9] Spanier [11].

In these cohomology groups we drop the coefficient group to simplify our notation, whenever there is no danger of confusion. The case that the subset D or both D and B are empty is not excluded. When several inclusion mappings are under consideration, we shall denote them also by i, k, or l, or distinguish them by subscripts. These induced homomorphisms have some simple properties, which have been taken as axioms by Eilenberg and Steenrod in their axiomatic treatment of homology theory.[10] The following axioms will be frequently used in our discussions:

1. The excision axiom.[11] If A, B are closed subsets of X, the homomorphism

$$j^*\colon H^r(A \cup B, B) \to H^r(A, A \cap B) \tag{18}$$

is an onto isomorphism.

2. The exactness axiom. Let

$$\delta^*\colon H^{r-1}(A) \to H^r(X, A) \tag{19}$$

be the coboundary homomorphism. Then the sequence

$$\begin{aligned} \cdots \to H^{r-1}(A) &\xrightarrow{\delta^*} H^r(X, A) \xrightarrow{j^*} H^r(X) \\ &\xrightarrow{i^*} H^r(A) \xrightarrow{\delta^*} H^{r+1}(X, A) \to \cdots \end{aligned} \tag{20}$$

is exact.

Now let A, B be closed subsets of X, such that $B \subset A \subset X$. Consider the sequence

$$H^r(A, B) \xrightarrow{j^*} H^r(A) \xrightarrow{\delta^*} H^{r+1}(X, A),$$

and define the homomorphism $\Delta^* = \delta^* j^*$. It follows from (20) that the sequence

$$\begin{aligned} \cdots \to H^r(X, A) &\xrightarrow{j^*} H^r(X, B) \\ &\xrightarrow{k^*} H^r(A, B) \xrightarrow{\Delta^*} H^{r+1}(X, A) \to \cdots \end{aligned} \tag{21}$$

is exact. It is called the *exact sequence of a triple* $B \subset A \subset X$.

---

[10] Eilenberg and Steenrod [5].

[11] This is a strengthened form of the excision axiom and is not true for general homology theory. It is true for homology theories invariant under what Eilenberg and Steenrod called relative homeomorphisms. An example is given by the Cech homology or cohomology theory for the category of compact pairs (Cf. [5], 266, Theorem 5.4). In our applications all the spaces under consideration are finite polyhedra, for which this excision axiom is valid.

LEMMA 1.1. *Let X be a topological space, and A, B, C closed subsets, such that*

$$X = A \cup B, \qquad A \cap B \subset C.$$

Then we have the isomorphism

$$H^r(X, C) \cong H^r(A, A \cap C) \oplus H^r(B, B \cap C). \tag{22}$$

To prove this, we consider the following groups related by homomorphisms, all induced by inclusion mappings:

$$\begin{tikzcd} H^r(X, B \cup C) \arrow[r, "j^*"] & H^r(X, C) \arrow[r, "k^*"] & H^r(B \cup C, C) \\ & H^r(X, A \cup C) \arrow[u, "i^*"] \arrow[ur, "l^*"'] & \end{tikzcd}$$

The groups of the first row are taken from the exact sequence of the triple $C \subset B \cup C \subset X$, and therefore form an exact sequence. By the excision axiom, l^* is an onto isomorphism. Writing $i^* l^{*-1} = \lambda$, we have $k^* \lambda =$ identity. For $x \in H^r(X, C)$, we find $k^*(x - \lambda k^* x) = 0$, which allows us to put $x - \lambda k^* x = j^* y$, $y \in H^r(X, B \cup C)$. On the other hand, if $x = \lambda z = j^* y$, $z \in H^r(B \cup C, C)$, then $z = k^* \lambda z = k^* j^* y = 0$. It follows that $H^r(X, C)$ is a direct sum of $j^* H^r(X, B \cup C)$ and $\lambda H^r(B \cup C, C)$. λ is clearly an isomorphism (into), so that i^* is an isomorphism. By symmetry between A and B it follows that j^* is an isomorphism. Since, by the excision axiom, the homomorphisms

$$H^r(X, A \cup C) \to H^r(B, B \cap C),$$
$$H^r(X, B \cup C) \to H^r(A, A \cap C),$$

induced by the inclusion mappings are onto isomorphisms, the lemma follows.

By induction this lemma can be put in the following generalized form:

LEMMA 1.1′. *Let X be a topological space, and $A_1, \cdots, A_s$, C closed subsets, such that*

$$X = A_1 \cup \cdots \cup A_s, \qquad A_i \cap A_k \subset C, \qquad i \neq k;\ i, k = 1, \cdots, s.$$

Then

$$H^r(X, C) \cong \sum_{i=1}^{s} H^r(A_i, A_i \cap C). \tag{22′}$$

the right-hand side being a direct sum of groups.

We now consider a fiber bundle $p: B \to X$, whose base space X is a

finite connected complex of dimension n and whose fiber F is connected. Denote by X^k the k-dimensional skeleton of X, and put $B_k = p^{-1}(X^k)$. It will turn out that the relative cohomology groups $H^r(B_q, B_{q-1}; G)$ can be interpreted in a simple manner. For simplicity we shall suppose our coefficient system to be simple, an assumption which is fulfilled in all our later applications.

LEMMA 1.2. *The group $H^r(B_q, B_{q-1}; G)$ is isomorphic to the direct sum*

$$\sum H^r(p^{-1}(\bar{\sigma}), p^{-1}(\dot{\sigma}); G) \cong C^q(X; H^{r-q}(F; G)), \tag{23}$$

where the summation is over all the q-dimensional cells of X and where the group $C^q(X; H^{r-q}(F; G))$ is the group of all q-dimensional cochains of X with the coefficient group $H^{r-q}(F; G)$. If λ denotes this isomorphism of $H^r(B_q, B_{q-1}; G)$ onto $C^q(X; H^{r-q}(F; G))$, then commutativity holds in the diagram

$$\begin{array}{ccccc} H^r(B_q, B_{q-1}; G) & \xrightarrow{j^*} & H^r(B_q; G) & \xrightarrow{\delta^*} & H^{r+1}(B_{q+1}, B_q; G) \\ \lambda \downarrow & & & & \downarrow \lambda \\ C^q(X; H^{r-q}(F; G)) & & \xrightarrow{\quad\delta\quad} & & C^{q+1}(X; H^{r-q}(F; G)), \end{array} \tag{24}$$

that is, $\delta\lambda = \lambda\delta^* j^*$, where δ is the coboundary operator for the group of cochains of X.

We denote by $\bar{\sigma}_a{}^q$, $a = 1, \cdots, s$, the closed q-cells of X, by $\dot{\sigma}_a{}^q$ the set-theoretical boundary of $\bar{\sigma}_a{}^q$, and write $\sigma_a{}^q = \bar{\sigma}_a{}^q - \dot{\sigma}_a{}^q$. The latter will also denote the corresponding cell, chain, or cochain, when it is oriented. Putting $A_a = p^{-1}(\bar{\sigma}_a{}^q)$, $C = B_{q-1}$, we get, by applying Lemma 1.1′ to the space B_q, the direct sum decomposition

$$H^r(B_q, B_{q-1}; G) \cong \sum_{a=1}^{s} H^r(p^{-1}(\bar{\sigma}_a{}^q), p^{-1}(\dot{\sigma}_a{}^q); G). \tag{23′}$$

To describe this isomorphism more explicitly, we put, for every a, $B_{q,a} = p^{-1}(X^q - \sigma_a{}^q)$, and consider the following homomorphisms, all induced by inclusion mappings:

$$\begin{tikzcd} & H^r(B_q, B_{q,a}; G) \arrow[dl, "i_a^*"'] \arrow[dr, "j_a^*"] & \\ H^r(p^{-1}(\bar{\sigma}_a{}^q), p^{-1}(\dot{\sigma}_a{}^q); G) & & H^r(B_q, B_{q-1}; G). \arrow[ll, "k_a^*"'] \end{tikzcd}$$

By the excision axiom, $i_a{}^*$ is an onto isomorphism, while, by Lemma 1.1,

j_a^* is an isomorphism. Moreover, we have $i_a^* = k_a^* j_a^*$ or $k_a^*(j_a^* i_a^{*-1}) =$ identity. In the decomposition (23′), $H^r(B_q, B_{q-1}; G)$ is a direct sum of the subgroups $j_a^* i_a^{*-1} H^r(p^{-1}(\bar{\sigma}_a^q), p^{-1}(\dot{\sigma}_a^q); G)$.

Since $p^{-1}(\bar{\sigma}_a^q)$ is homeomorphic to $\bar{\sigma}_a^q \times F$, we have

$$H^r(p^{-1}(\bar{\sigma}_a^q), p^{-1}(\dot{\sigma}_a^q); G) \cong H^{r-q}(F; G) \otimes H^q(\bar{\sigma}_a^q, \dot{\sigma}_a^q; Z),$$

where Z is the additive group of integers. The latter group is isomorphic to the group of cochains $\alpha\sigma_a^q$, $\alpha \varepsilon H^{r-q}(F; G)$. If μ_a denotes this isomorphism, λ is defined componentwise by $\lambda_a = \mu_a k_a^*$.

To prove the commutativity of the diagram (24), it is sufficient to take $x \varepsilon H^r(B_q, B_{q,a}; G)$ and to prove that $\delta\lambda j_a^*(x)$ and $\lambda\delta^* j^* j_a^*(x)$, both $(q+1)$-dimensional cochains of X, have the same value for any $(q+1)$-cell σ_b^{q+1}. For this purpose it is essential to consider the following diagram:

$$\begin{tikzcd}[column sep=tiny]
 & H^r(B_q, B_{q,a}; G) \arrow[dl, "i_a^*"'] \arrow[d, "l_{ab}^*"] \arrow[r, "\Delta^*"] & H^{r+1}(B_{q+1}, B_q; G) \arrow[dr, "k_b^*"] & \\
H^r(p^{-1}(\bar{\sigma}_a^q), p^{-1}(\dot{\sigma}_a^q); G) \arrow[dr, "\mu_a"'] & H^r(p^{-1}(\dot{\sigma}_b^{q+1}), p^{-1}(\dot{\sigma}_b^{q+1} - \sigma_a^q); G) \arrow[r, "\Delta_1^*"] & H^{r+1}(p^{-1}(\bar{\sigma}_b^{q+1}), p^{-1}(\dot{\sigma}_b^{q+1}); G) \arrow[dl, "\mu_b"] \\
 & C^q(\bar{\sigma}_a^q; H^{r-q}(F; G)) \arrow[r, "\delta"] & C^{q+1}(\bar{\sigma}_b^{q+1}; H^{r-q}(F; G)). &
\end{tikzcd}$$

In this diagram i_a^*, k_b^*, and l_{ab}^* are induced by inclusion mappings, δ is the coboundary operator of the group of cochains, while $\Delta^* = \delta^* j^* j_a^*$ and Δ_1^* are the coboundary operators of the triples $B_{q,a} \subset B_q \subset B_{q+1}$ and $p^{-1}(\dot{\sigma}_b^{q+1} - \sigma_a^q) \subset p^{-1}(\dot{\sigma}_b^{q+1}) \subset p^{-1}(\bar{\sigma}_b^{q+1})$. In the last notation we adopt the convention that $\dot{\sigma}_b^{q+1} - \sigma_a^q = \dot{\sigma}_b^{q+1}$, if σ_a^q is not a face of σ_b^{q+1}. Since the second triple can be mapped into the first one by the inclusion mapping, we have $k_b^* \Delta^* = \Delta_1^* l_{ab}^*$. Since we are only interested in the values of the cochains for σ_b^{q+1}, we can restrict ourselves to the bundle over $\bar{\sigma}_b^{q+1}$. Then we have

$$\lambda\Delta^* = \mu_b k_b^* \Delta^* = \mu_b \Delta_1^* l_{ab}^*,$$

$$\delta\lambda j_a^* = \delta\mu_a k_a^* j_a^* = \delta\mu_a i_a^*.$$

It suffices to prove that the homomorphisms in the right-hand members of these two equations are identical.

If σ_a^q is not a face of σ_b^{q+1}, both homomorphisms will give zero, because $H^r(p^{-1}(\dot{\sigma}_b^{q+1}), p^{-1}(\dot{\sigma}_b^{q+1} - \sigma_a^q); G) = 0$ and δ obviously gives zero. Suppose now σ_a^q be a face of σ_b^{q+1}. By the excision axiom, the homomorphism

$$i_{ab}^*: H^r(p^{-1}(\dot{\sigma}_b^{q+1}), p^{-1}(\dot{\sigma}_b^{q+1} - \sigma_a^q); G) \to H^r(p^{-1}(\bar{\sigma}_a^q), p^{-1}(\dot{\sigma}_a^q); G)$$

induced by the inclusion mapping is an onto isomorphism. Moreover, $i_a^* = i_{ab}^* l_{ab}^*$. It suffices therefore to prove that $\mu_b \Delta_1^* = \delta \mu_a i_{ab}^*$. But then all the groups in question refer to the cell $\bar{\sigma}_b^{q+1}$ or to the bundle over it, which is homeomorphic to the Cartesian product $\bar{\sigma}_b^{q+1} \times F$. The verification of the relation in this case is trivial. This completes the proof of Lemma 1. 2.

Lemma 1. 2 can be briefly described by saying that the relative cohomology group $H^r(B_q, B_{q-1}; G)$ is isomorphic to the group of q-dimensional cochains of X with the coefficient group $H^{r-q}(F; G)$ and that the homomorphism $\delta^* j^*$ becomes then the coboundary operator under this identification. Because of our assumption that the coefficient system is simple, consideration of local coefficients is not necessary.

However, we are interested not in the relative cohomology groups $H^r(B_q, B_{q-1}; G)$, but in the absolute cohomology groups $H^r(B; G) = H^r(B_n; G)$. To derive information about them, we consider successively the groups $H^r(B_q; G)$, $q = 0, 1, \cdots, n$. All these groups are connected by homomorphisms as in the diagram:

$$
\begin{array}{ccccc}
H^r(B_n, B_{n-1}) & \xrightarrow{j^*} & H^r(B_n = B) & & \\
 & & \downarrow i^* & & \\
 & & \vdots & & \\
 & & \downarrow i^* & & \\
\cdots \to H^r(B_q, B_{q-1}) & \xrightarrow{j^*} & H^r(B_q) & \xrightarrow{\delta^*} & H^{r+1}(B_{q+1}, B_q) \to \cdots \\
 & & \downarrow i^* & & \\
 & & \vdots & & \\
 & & \downarrow i^* & & \\
\cdots \to H^r(B_r, B_{r-1}) & \xrightarrow{j^*} & H^r(B_r) & \xrightarrow{\delta^*} & H^{r+1}(B_{r+1}, B_r) \to \cdots \\
 & & \downarrow i^* & & \\
 & & \vdots & & \\
 & & \downarrow i^* & & \\
H^r(B_0, B_{-1}) & \xrightarrow{j^*} & H^r(B_0) & \xrightarrow{\delta^*} & H^{r+1}(B_1, B_0). \\
 & & \downarrow & & \\
 & & 0 & &
\end{array}
\tag{25}
$$

In all these cohomology groups the coefficient group is G, which is omitted for simplicity of notation. Every sequence of groups from the diagram connected by homomorphisms in the cyclic order j^*, i^*, δ^* is exact. We shall denote by $H_0^r(B_q)$ that subgroup of $H^r(B_q)$, which is the kernel of δ^*.

First a remark about the homology position of the fiber in the bundle B. We take a vertex $v \varepsilon X^0$, and identify F with $p^{-1}(v)$. This gives rise to the

inclusion mappings $l_0\colon F \to B_0$ and $l\colon F \to B$, and the induced homomorphisms

$$\begin{tikzcd} & H^r(B) \arrow[dl, "l^*"'] \arrow[d, "i^*"] \\ H^r(F) & H_0{}^r(B_0) \arrow[l, "l_0^*"'] \end{tikzcd} \tag{26}$$

The homomorphism l_0^* in (26) is an onto isomorphism, for $H_0{}^r(B_0)$ is clearly isomorphic to the group of 0-dimensional cohomology classes of X, with the coefficient group $H^r(F)$, which is isomorphic to $H^r(F)$.

The diagram (25) leads to a homological definition of the primary obstruction, as given by the lemma:

LEMMA 1.3. *Let $r > 1$ be the integer such that $\pi_r(F) \neq 0$ and $\pi_s(F) = 0$ for all $s < r$. Let $H_r(F; Z)$ be the coefficient group G and $\omega \in H^r(F; G)$ be the cohomology class which assigns to any $z \in H_r(F; Z)$ the element z itself. There exists $\bar{\omega} \in H^r(B_r; G)$ such that $(i^*)^r\bar{\omega} = l_0^{*-1}\omega$; the element $\lambda\delta^*\bar{\omega}$ is a cocycle and its cohomology class is the negative of the primary obstruction.*

Consider first an element $\phi \in H_0{}^r(B_0)$. From the relations

$$H^{r+1}(B_s, B_{s-1}) = 0, \qquad 1 < s < r+1,$$

it follows that there exists $\bar{\phi} \in H^r(B_r)$ such that $(i^*)^r\bar{\phi} = \phi$. By Lemma 1.2 we have

$$\delta(\lambda\delta^*\bar{\phi}) = \lambda\delta^*j^*\delta^*\bar{\phi} = 0,$$

which means that $\lambda\delta^*\bar{\phi}$ is a cocycle. On the other hand, $\bar{\phi}$ is determined up to an additive term j^*y, $y \in H^r(B_r, B_{r-1})$, so that $\lambda\delta^*\bar{\phi}$ is determined up to an additive term $\lambda\delta^*j^*y = \delta\lambda y$, that is, up to a coboundary. Thus the cohomology class of $\lambda\delta^*\bar{\phi}$ is completely determined.

We also remark that under our assumptions the natural homomorphism

$$f\colon H^r(F; G) \to \operatorname{Hom}(G, G), \qquad G = H_r(F; Z), \tag{27}$$

is an onto isomorphism.[12] The class ω in the statement of the Lemma is therefore well defined. From the above it follows that the same is true of the cohomology class of $\lambda\delta^*\bar{\omega}$. The verification that this is equal to the

[12] Eilenberg-MacLane [4], 808, Theorem 32.1. The theorem quoted was formulated for a star-finite complex and for homology groups with infinite cycles. A similar statement is therefore true for a closure-finite complex and for cohomology groups with infinite cocycles. Our result follows from this theorem and the further fact that $H^{r-1}(F; G) = 0$.

negative of the primary obstruction can be carried out by studying a cross-section over X^r. It is straightforward and we shall omit the details here.

We are now in a position to give a proof of Theorem 1 (Cf. Introduction). Since the primary obstruction is zero, the element $\bar{\omega} \varepsilon H^r(B_r)$ can be so chosen that $\delta^* \bar{\omega} = 0$. Since

$$H^{r+1}(B_s, B_{s-1}) = 0, \qquad s > r + 1,$$

there is $\bar{\bar{\omega}} \varepsilon H^r(B)$ such that $(i^*)^{n-r} \bar{\bar{\omega}} = \bar{\omega}$, which implies

$$(i^*)^n \bar{\bar{\omega}} = (i^*)^r \bar{\omega} = l_0^{*-1} \omega.$$

By the definition of our notation for inclusion mappings, we can write i^* for $(i^*)^n$, and the last relation becomes $l_0^* i^* \bar{\bar{\omega}} = l^* \bar{\bar{\omega}} = \omega$. Thus ω belongs to the image of l^*.

To prove that the same is true of any element $\phi \varepsilon H^r(F; G)$, we consider its corresponding endomorphism $f(\phi)$ of G into itself. This endomorphism $f(\phi)$ of the coefficient group induces endomorphisms ϕ_B and ϕ_F of $H^r(B; G)$ and $H^r(F; G)$ respectively. Moreover, it is clear that $\phi_F(\omega) = \phi$. From the commutativity of the diagram

$$\begin{array}{ccc} H^r(B;G) & \xrightarrow{\phi_B} & H^r(B;G) \\ l^* \downarrow & & \downarrow l^* \\ H^r(F;G) & \xrightarrow{\phi_F} & H^r(F;G) \end{array}$$

it follows that ϕ belongs to the image of l^*. This completes the proof of Theorem 1.

The following Lemma describes the operation of "integration over the fiber":

LEMMA 1.4. *Let $H^r(F;Z) \neq 0$, $r > 0$, be the last non-vanishing cohomology group of F, so that $H^s(F;Z) = 0$ for all $s > r$. To any coefficient group G an integration over the fiber*

$$(1) \qquad \Psi \colon H^m(B;G) \to H^{m-r}(X; H^r(F;G)),$$

can be defined.

This homomorphism Ψ has been described in the Introduction. To prove its existence we first notice that our assumptions imply

$$H^{m-1}(B_{m-r-2}; G) = H^m(B_{m-r-1}; G) = 0.$$

Let $u \varepsilon H^m(B)$ and $u' = (i^*)^{m-r} u \varepsilon H^m(B_{m-r})$ and consider the cohomology groups

$$\begin{array}{ccccccc}
H^{m-1}(B_{m-r-1}, B_{m-r-2}) & \xrightarrow{j^*} & H^{m-1}(B_{m-r-1}) & & & & \\
& & i^* \downarrow & & & & \\
& & 0 & & & & \\
& & & & & & \\
& \xrightarrow{\delta^*} & H^m(B_{m-r}, B_{m-r-1}) & \xrightarrow{j^*} & H^m(B_{m-r}) & \xrightarrow{\delta^*} & H^{m+1}(B_{m-r+1}, B_{m-r}) \\
& & & & i^* \downarrow & & \\
& & & & 0 & &
\end{array}$$

where the sequences in the cyclic order j^*, i^*, δ^* are exact. Since $i^*u' = 0$, there is $v \varepsilon H^m(B_{m-r}, B_{m-r-1})$ with $u' = j^*v$. Then $\lambda v \varepsilon C^{m-r}(X; H^r(F; G))$ is a cocycle, for $\delta\lambda v = \lambda\delta^*j^*v = \lambda\delta^*u' = 0$. Moreover, v is defined up to an additive term δ^*y, $y \varepsilon H^{m-1}(B_{m-r-1})$, where $y = j^*z$, $z \varepsilon H^{m-1}(B_{m-r-1}, B_{m-r-2})$. Since $\lambda\delta^*y = \lambda\delta^*j^*z = \delta\lambda z$, λv is defined up to a coboundary. Its cohomology class, which is an element of $H^{m-r}(X; H^r(F; G))$, is therefore completely determined by u. We call it Ψu and thus prove the existence of the homomorphism Ψ.

Concerning the homomorphism Ψ, we have the following useful lemma which follows immediately from its definition:

LEMMA 1.5. *Let $p\colon B \to X$ and $p'\colon B' \to X'$ be two fiber bundles and $\tilde{f}\colon B \to B'$ be a bundle map which induces a mapping $f\colon X \to X'$ of the base spaces. If Ψ, Ψ' denote integrations over the fiber of the two bundles, then commutativity holds in the diagram:*

$$\begin{array}{ccc}
H^m(B) & \xleftarrow{\quad \tilde{f}^* \quad} & H^m(B') \\
\Psi \downarrow & & \downarrow \Psi' \\
H^{m-r}(X; H^r(F)) & \xleftarrow{f^*} & H^{m-r}(X'; H^r(F)),
\end{array}$$

that is, $f^*\Psi' = \Psi\tilde{f}^*$.

2. Proof of Theorem 2 (the Generalized Leray-Hirsch Theorem).

To carry out the proof we shall adopt the notations of the preceding section. By double induction on r and q, we proceed to prove the following statements:

a) There exists a homomorphism $\mu_{rq}\colon H_0{}^r(B_q) \to H^r(B)$, such that $(i^*)^{n-q}\mu_{rq}$ is the identity automorphism of $H_0{}^r(B_q)$;

b) $H_0{}^r(B_q)$ is isomorphic to the direct sum of $H_0{}^r(B_{q-1})$ and $H^q(X; H^{r-q}(F))$. Here we make the convention that a cohomology group of negative dimension is vacuous.

We remark that, for $q = n$, b) implies that $H^r(B)$ is isomorphic to the

direct sum $\sum_{s=0}^{n} H^s(X; H^{r-s}(F))$. Since the latter is isomorphic to $H^r(X \times F)$ by Künneth's Theorem, our theorem follows from b).

For $q = 0$ we put $\mu_{r0} = \mu l_0^*$. Since l_0^* establishes an isomorphism between $H_0^r(B_0)$ and $H^r(F)$, the fact that $l^*\mu = l_0^* i^* \mu$ is the identity automorphism of $H^r(F)$ implies that $l_0^{*-1}(l_0^* i^* \mu) l_0^* = i^* \mu_{r0}$ is the identity automorphism of $H_0^r(B_0)$. This proves a). Statement b) is obvious, if we define $H_0^r(B_{-1})$ to be zero.

Suppose μ_{ts} be defined for $t < r$ and $t = r$, $0 \leqq s \leqq q-1$, fulfilling the conditions a), b). Consider the diagram

$$\begin{tikzcd}
\arrow[r] & H^{r-1}(B_{q-1}, B_{q-2}) \arrow[r, "j^*"] & H^{r-1}(B_{q-1}) \arrow[r, "\delta^*"] \arrow[d, "i^*"] & H^r(B_q, B_{q-1}) \arrow[d, "j^*"] & \\
& & H^{r-1}(B_{q-2}) & H^r(B_q) \arrow[r, "\delta^*"] \arrow[d, "i^*"] & H^{r+1}(B_{q+1}, B_q) \arrow[r] & \cdots . \\
& & & H^r(B_{q-1}) &
\end{tikzcd}$$

We put $\nu = (i^*)^{n-q} \mu_{r,q-1}\colon H_0^r(B_{q-1}) \to H_0^r(B_q)$, so that $i^*\nu =$ identity by induction hypothesis. A familiar argument proves that $H^r(B_q)$ is a direct sum of $\nu H_0^r(B_{q-1})$ and $j^* H^r(B_q, B_{q-1})$. Moreover, ν is clearly an isomorphism. By exactness the second summand fulfills the isomorphism

$$j^* H^r(B_q, B_{q-1}) \cong H^r(B_q, B_{q-1}) / \delta^* H^{r-1}(B_{q-1}).$$

From our induction hypothesis it follows that the group $H^{r-1}(B_{q-1})$ is a direct sum of $(i^*)^{n-q+1} \mu_{r-1,q-2} H_0^{r-1}(B_{q-2})$ and $j^* H^{r-1}(B_{q-1}, B_{q-2})$ of which the first summand goes to 0 under δ^*. Therefore we have

$$j^* H^r(B_q, B_{q-1}) \cap H_0^r(B_q) \cong K^r(B_q, B_{q-1}) / \delta^* j^* H^{r-1}(B_{q-1}, B_{q-2}),$$

where $K^r(B_q, B_{q-1}) \subset H^r(B_q, B_{q-1})$ is the kernel of $\delta^* j^*$. Using the isomorphism λ, we see that this group is isomorphic to $H^q(X; H^{r-q}(F))$. This proves b).

It remains to define μ_{rq} to satisfy a). For an element

$$x = \nu y \,\varepsilon\, \nu H_0^r(B_{q-1}),\ y \,\varepsilon\, H_0^r(B_{q-1}),$$

we set

$$\mu_{rq} x = \mu_{r,q-1} y = \mu_{r,q-1} i^* x.$$

Then $(i^*)^{n-q} \mu_{rq} = (i^*)^{n-q} \mu_{r,q-1} i^* = \nu i^*$ is the identity automorphism, since ν is an isomorphism of $H_0^r(B_{q-1})$ onto $\nu H_0^r(B_{q-1})$. To define μ_{rq} for the other summand, we need some preparations.

First we put $B' = X \times B$ and consider B' as a bundle over X with projection p' defined by $p'(x, b) = x$, $x \varepsilon X, b \varepsilon B$. To this bundle the homology theory established in § 1 can be applied, and we shall denote the notions and symbols pertaining to it by dashes. This bundle is, however, a trivial bundle and has very simple properties. In particular, we have, by Künneth's Theorem, the isomorphisms

$$H_0{}^r(B_q') = H_0{}^r(X^q \times B) \cong \sum_{s=0}^{q} H^s(X; H^{r-s}(B)),$$

$$H^r(B') = H^r(X \times B) \cong \sum_{s=0}^{n} H^s(X; H^{r-s}(B)).$$

These permit us to define the homomorphisms

$$\mu_{rq}': H_0{}^r(B_q') \to H^r(B'),$$

such that $(i'^*)^{n-q}\mu_{rq}'$ is the identity.

Next we define the mappings

$$g_q: B_q \to B_q' = X^q \times B$$

by $g_q(b) = (p(b), b)$, $b \varepsilon B_q$, and write $g_n = g$. These mappings induce homomorphisms on the cohomology groups, for which there is commutativity in the diagram

$$\begin{array}{ccccc}
H^r(X^q \times B, X^{q-1} \times B) & \xrightarrow{j'^*} & H^r(X^q \times B) & \xrightarrow{\delta'^*} & H^{r+1}(X^{q+1} \times B, X^q \times B) \\
\downarrow g_q{}^* & & \downarrow g_q{}^* & & \downarrow g_{q+1}{}^* \\
H^r(B_q, B_{q-1}) & \xrightarrow{j^*} & H^r(B_q) & \xrightarrow{\delta^*} & H^{r+1}(B_{q+1}, B_q).
\end{array}$$

To $y' \varepsilon H^r(X^q \times B, X^{q-1} \times B)$, $y \varepsilon H^r(B_q, B_{q-1})$, we have $\lambda' y' \varepsilon C^q(X; H^{r-q}(B))$, $\lambda y \varepsilon C^q(X; H^{r-q}(F))$, so that we can write

$$\lambda' y' = \sum_i h_i' \sigma_i{}^q, \qquad h_i' \varepsilon H^{r-q}(B),$$

$$\lambda y = \sum_i h_i \sigma_i{}^q, \qquad h_i \varepsilon H^{r-q}(F),$$

where $\sigma_i{}^q$ are the q-cells of X. It follows from the definition of g_q that

$$g_q{}^* y' = \lambda^{-1}\{\sum_i l^*(h_i')\sigma_i{}^q\}.$$

Conversely, because of the existence of the homomorphism μ, we can define the homomorphism

$$\rho_{rq}: H^r(B_q, B_{q-1}) \to H^r(X^q \times B, X^{q-1} \times B)$$

by the equation

$$\rho_{rq}(y) = \lambda'^{-1}\{\sum \mu(h_i)\sigma_i{}^q\}.$$

It has therefore the property that $g_q{}^*\rho_{rq} = \text{identity}$. Moreover, from the interpretation of $\delta^* j^*$ and $\delta'^* j'^*$ as coboundary operators under the isomorphisms λ and λ', we have

$$\rho_{r+1,q+1}\delta^* j^* = \delta'^* j'^* \rho_{rq}.$$

From this it follows that $\delta^* j^* y = 0$ if and only if $\delta'^* j'^* \rho_{rq} y = 0$.

To define μ_{rq} for the summand $j^* H^r(B_q, B_{q-1}) \cap H_0{}^r(B_q)$ of $H_0{}^r(B_q)$, we take $y \varepsilon H^r(B_q, B_{q-1})$ with $\delta^* j^* y = 0$. Then $j'^* \rho_{rq} y \varepsilon H_0{}^r(B'_q)$ and $\mu_{rq}' j'^* \rho_{rq} y \varepsilon H^r(B') = H^r(X \times B)$. Now consider the diagram

$$\begin{array}{ccc} H^r(X \times B) & \xrightarrow{g^*} & H^r(B), \\ \downarrow (i'^*)^{n-q} & & \downarrow (i^*)^{n-q} \\ H^r(X^q \times B) & \xrightarrow{g_q{}^*} & H^r(B_q) \end{array}$$

in which commutativity holds. We define

$$\mu_{rq} j^* y = g^* \mu_{rq}' j'^* \rho_{rq} y.$$

When we modify y by an additive term $\delta^* j^* z$, $z \varepsilon H^{r-1}(B_{q-1}, B_{q-2})$, the right-hand side will be modified by an additive term

$$g^* \mu_{rq}' j'^* \rho_{rq} \delta^* j^* z = g^* \mu_{rq}' j'^* \delta'^* j'^* \rho_{r-1,q-1} z = 0.$$

Hence μ_{rq} depends only on $j^* y$ and not on the choice of y. Since

$$\begin{aligned} (i^*)^{n-q} \mu_{rq} j^* y &= (i^*)^{n-q} g^* \mu_{rq}' j'^* \rho_{rq} y = g_q{}^* (i'^*)^{n-q} \mu_{rq}' j'^* \rho_{rq} y \\ &= g_q{}^* j'^* \rho_{rq} y = j^* g_q{}^* \rho_{rq} y = j^* y, \end{aligned}$$

we conclude that the homomorphism μ_{rq} satisfies condition a).

This completes our induction and hence the proof of Theorem 2.

The following corollary follows immediately from the above proof.

COROLLARY 2. 1. *Suppose the hypotheses of Theorem 2 be satisfied. Then* $H^m(B)$ *is isomorphic to the direct sum* $\sum_{q=0}^{m} H^q(X; H^{m-q}(F))$. *For* $u \varepsilon H^m(B)$, *its image* Ψu *under the integration over the fiber* Ψ *is the component of* u *in the summand* $H^{m-r}(X; H^r(F))$, *where* r *is defined by Lemma* 1. 4.

Under some further assumptions which will be satisfied in our applications, we can express these relations in a more explicit form, using our homomorphism μ. In fact, from our proof of Theorem 2, we see that $H^r(B)$ is a direct sum of $\mu_{rq}(j^* H^r(B_q, B_{q-1}) \cap H_0{}^r(B_q))$ for $q = 0, 1, \cdots, r$.

Suppose that our coefficient system is a ring R with unit element. We call an element of $Z^q(X;H^{r-q}(F;R))$ a cocycle of the first kind, if it is equal to a finite sum of the form $\sum_i c_i \otimes z_i$, where $c_i \varepsilon Z^q(X;R)$, $z_i \varepsilon H^{r-q}(F;R)$. If this is not the case, we call it a cocycle of the second kind. We now make the assumption that every element of $H^q(X;H^{r-q}(F;R))$ has as representative a cocycle of the first kind. This condition is satisfied when, for instance, R is a field.

To an element $y \varepsilon H^r(B_q, B_{q-1})$, with $\delta^* j^* y = 0$, we can write

$$\lambda y = \sum_i c_i \otimes z_i,$$

where $c_i \varepsilon Z^q(X;R)$, $z_i \varepsilon H^{r-q}(F;R)$. The homomorphism μ_{rq}' can be defined such that

$$\mu_{rq}' j'^* \rho_{rq} y = \sum_i \gamma_i \otimes \mu(z_i),$$

where γ_i is the class of c_i. By definition, $\mu_{rq} j^* y$ is the image of this element under the homomorphism g^*. Now we can decompose g as a product of two mappings: $g = h\Delta$, where $\Delta : B \to B \times B$ is the diagonal map defined by $\Delta(b) = (b, b)$, $b \varepsilon B$, and h: $B \times B \to X \times B$ is defined by $h(b, b') = (p(b), b')$, $b, b' \varepsilon B$. It follows that

$$g^*(\sum_i \gamma_i \otimes \mu(z_i)) = \Delta^* h^*(\sum_i \gamma_i \otimes \mu(z_i)) = \Delta^*(\sum_i p^*(\gamma_i) \otimes \mu(z_i)) = \sum_i p^*(\gamma_i) \cup \mu(z$$

These considerations lead to the following theorem:

Corollary 2.2. *Under the hypotheses of Theorem 2 suppose further that, for an integer $m \geqq 0$, every element of $H^q(X;H^{m-q}(F;R))$ has as representative a cocycle of the first kind. Then every element $u \varepsilon H^m(B;R)$ can be written in the form*

$$u = \sum_i \mu(z_i) \cup p^*(\gamma_i), \qquad z_i \varepsilon H^q(F;R), \gamma_i \varepsilon H^{m-q}(X;R),$$

where $\dim z_i + \dim \gamma_i = m$. *If r is the integer such that $H^r(F;R)$ is the highest non-vanishing cohomology group of F with the coefficient group R, then*

$$\Psi u = \sum_{\dim z_i = r} z_i \otimes \gamma_i.$$

3. The case of manifolds; proof of Theorem 3. Throughout this section we suppose that B, X, F are oriented manifolds, with the convention that these same symbols denote their fundamental homology and cohomology classes. It is easily seen that $\Psi(B) = \pm X$. We supose that the orientations

of the manifolds are such that the positive sign holds. Under our assumptions the highest non-vanishing cohomology group of F is $H^s(F; G)$, where s is the dimension of F, and it is naturally isomorphic to G. Integration over the fiber can therefore be considered as a homomorphism

$$\Psi\colon H^m(B; G) \to H^{m-s}(X; G). \tag{28}$$

LEMMA 3. 1. *Let the abelian groups G_1 and G_2 be paired to G, and let $v \in H^q(X; G_1)$, $u \in H^r(B; G_2)$. Then*

$$\Psi(p^*v \cup u) = v \cup \Psi u. \tag{29}$$

To prove this lemma we denote by $d\colon X \to X \times X$ the diagonal map defined by $d(x) = (x, x)$, $x \in X$. Similarly, let $\Delta\colon B \to B \times B$ be the diagonal map of B. We define the mappings

$$p'\colon B \times B \to X \times B, \tag{30}$$

$$p''\colon X \times B \to X \times X, \tag{31}$$

by $p'(b, b') = (p(b), b')$, $p''(x, b) = (x, p(b))$, $b, b' \in B$, $x \in X$. These give rise to two bundles, both with fibers F. We put $\Gamma = p'\Delta$. Then $\Gamma\colon B \to X \times B$ is a bundle map of the given bundle $p\colon B \to X$ into the bundle (31). By Lemma 1. 5, we have commutativity in the diagram $(m = q + r)$:

$$\begin{tikzcd}
 & H^m(B \times B; G) \arrow[dl, "\Delta^*"'] \\
H^m(B; G) \arrow[d, "\Psi"] & H^m(X \times B; G) \arrow[l, "\Gamma^*"'] \arrow[u, "p'^*"'] \arrow[d, "\Psi''"] \\
H^m(X; G) & H^m(X \times X; G) \arrow[l, "d^*"']
\end{tikzcd}$$

where Ψ'' is the integration over the fiber of the bundle (31).

Now let u', v' be representative cocycles of the cohomology classes u, v respectively. Then $v' \times u' \in Z^m(X \times B; G)$, the group of all m-cocycles of $X \times B$. If we denote by the same symbols p', Δ cellular approximations of the respective continuous mappings, we have

$$\Delta^*p'^*(v' \times u') = \Delta^*(p^*v' \times u') = p^*v' \cup u'.$$

This proves that the class containing $v' \times u'$ is mapped by $\Delta^*p'^*$ into $p^*v \cup u$. On the other hand, this class is mapped by Ψ'' into the class containing $v' \times \Psi u'$, which goes to $v \cup \Psi u$ under d^*. From the commutativity of our diagram follows therefore the formula (29).

We are now in a position to prove Theorem 3. As stated in the Theorem,

we have $u \varepsilon H^m(B;G)$. Then $\Psi u \varepsilon H^{m-s}(X;G)$ and $\mathfrak{D}_X \Psi u \varepsilon H_{n+s-m}(X;G)$. We take any element $v \varepsilon H^{n+s-m}(X; \mathrm{Char}\, G)$, and denote by a dot the multiplication of homology and cohomology classes. Then we have

$$v \cdot \mathfrak{D}_X \Psi u = v \cdot (\Psi u \cap X) = (v \cup \Psi u) \cdot X,$$
$$v \cdot p_* \mathfrak{D}_B u = (p^* v) \cdot (\mathfrak{D}_B u) = (p^* v) \cdot (u \cap B) = (p^* v \cup u) \cdot B$$
$$= \{\Psi(p^* v \cup u)\} \cdot X.$$

By Lemma 3.1, the right-hand sides of the two equations are equal. The same is therefore true of the left-hand sides, with arbitrary v. This proves Theorem 3.

4. On the bundles associated to a complex manifold. We now turn our attention to the study of a compact complex manifold M of (complex) dimension n. Some of the more important bundles associated to M have been described in the Introduction, and we shall follow the notations. This section will be devoted to the proof of a few elementary properties of these bundles.

LEMMA 4.1. *To $x \varepsilon M$ let $F_{nr} = q_{nr}^{-1}(x)$, $\tilde{F}_{nr} = (q_{nr} p_{nr})^{-1}(x)$ be the fibers of the bundles B_{nr}, $\tilde{B}_{nr}$ respectively. Then $p_{nr}(x) \colon \tilde{F}_{nr} \to F_{rn}$ is a bundle having as fiber the Cartesian product of r complex lines, each with the origin deleted. It gives rise to r characteristic cohomology classes $w_1, \cdots, w_r$ of dimension 2 in the base space F_{nr}, which generate, by ring operations, the cohomology ring of F_{nr} with integer coefficients and which can be so chosen that $w_1^{n-1} \cup \cdots \cup w_r^{n-r}$ is the fundamental cohomology class of F_{nr}.*

Let V_n be the tangent complex vector space at x, with 0 denoting its origin. A point of $\tilde{F}_{nr}$ is then a sequence of simple multivectors of V_n of the form $e_1, e_1 \wedge e_2, \cdots, e_1 \wedge \cdots \wedge e_r \neq 0$, while a point of F_{nr} is a sequence of linear spaces of directions $L_0 \subset L_1 \subset \cdots \subset L_{r-1}$. The projection $p_{nr}(x)$ assigns to the multivector $e_1 \wedge \cdots \wedge e_i$, $i = 1, \cdots, r$, the linear subspace L_{i-1} it determines. Since two simple multivectors determine the same L_{i-1} when and only when they differ from each other by a non-zero complex factor, the fiber of the bundle $p_{nr}(x) \colon \tilde{F}_{nr} \to F_{nr}$ is homeomorphic to a Cartesian product $E_1 \times \cdots \times E_r$ of r complex lines, each with the origin deleted. Moreover, we can suppose E_i to be the space of all $e_1 \wedge \cdots \wedge e_i$ having the same projection L_{i-1}. This bundle is orientable, since the spaces involved are complex manifolds. We denote by w_i the characteristic class corresponding to the factor E_i of the fiber. The classes $w_1, \cdots, w_r$ can be considered to be with integer coefficients.

To prove the assertions on the homology structure of F_{nr} we proceed by induction on r. For $r=1$, $\tilde{F}_{n1}=V_n-0$ and F_{n1} is the complex projective space P_{n-1} of dimension $n-1$. In this case we see easily that w_1 is the cohomology class dual to the homology class having as representative a hyperplane of P_{n-1}, so that the Lemma is true. Suppose that the Lemma holds for $r-1$. Let

$$t\colon F_{nr}\to F_{n,r-1} \tag{32}$$

be the mapping which sends the sequence $L_0\subset L_1\subset\cdots\subset L_{r-1}$ to $L_0\subset L_1\subset\cdots\subset L_{r-2}$. This defines a fiber bundle having as fiber the set of all L_{r-1} through a fixed L_{r-2} in P_{n-1}, which is homeomorphic to the complex projective space P_{n-r} of dimension $n-r$. Denote by $k_{nr}\colon P_{n-r}\to F_{nr}$ the inclusion mapping of a fiber into the bundle. When, in the bundle $\tilde{F}_{nr}\to F_{nr}$, we restrict the fiber to the product $E_1\times\cdots\times E_{r-1}$, it is induced by the mapping t. Hence the first $r-1$ characteristic classes in F_{nr} are $w_i=t^*w_i'$, $i=1,\cdots,r-1$, where w_i' are the characteristic classes in $F_{n,r-1}$ of the bundle $\tilde{F}_{n,r-1}\to F_{n,r-1}$. Let w_r be the class corresponding to the fiber E_r. Then $k_{nr}^*w_r=w$ is a generating class of P_{n-r}. We define $\mu(w^j)=w_r^j$, $j=1,\cdots,n-r-1$, and extend μ by linearity into a homomorphism of the cohomology groups of P_{n-r} into those of F_{nr} (with integer coefficients). Then $k_{nr}^*\mu=$ identity, and the hypotheses of Theorem 2 are satisfied. If Ψ is the integration over the fiber of the bundle (32), we see easily that

$$\Psi(w_1^{n-1}\cup\cdots\cup w_r^{n-r})=(w_1')^{n-1}\cup\cdots\cup(w_{r-1}')^{n-r+1}.$$

Since the right-hand side of this equation is the fundamental cohomology class of $F_{n,r-1}$ by our induction hypothesis, it follows that $w_1^{n-1}\cup\cdots\cup w_r^{n-r}$ is the fundamental cohomology class of F_{nr}. This completes the proof of the Lemma.

LEMMA 4.2. *The bundle B_{nr} satisfies the hypotheses of Theorem 2.*

As in the proof of Lemma 4.1, the bundle $\tilde{B}_{nr}\to B_{nr}$ has as fiber the Cartesian product of r complex lines, each with the origina deleted. Let $v_1,\cdots,v_r$ be the corresponding characteristic cohomology classes. Since the bundle $p_{nr}(x)\colon\tilde{F}_{nr}\to F_{nr}$ is induced by the inclusion mapping $l_{nr}\colon F_{nr}\to B_{nr}$, we have $l_{nr}^*v_i=w_i$, $i=1,\cdots,r$. Defining $\mu w_i^{\rho_i}=v_i^{\rho_i}$, $\rho_i=1,\cdots,n-i$, and extending it by linearity into a homomorphism of the cohomology groups of F_{nr} into those of B_{nr}, we have $l_{nr}^*\mu=$ identity. Hence the hypotheses of Theorem 2 are fulfilled.

5. Proof of the duality theorem. The proof of the duality theorem depends on the consideration of the universal bundles. Let $E_{\nu+N}$ be a complex vector space of dimension $\nu + N$ and let $H(\nu, N)$ be the Grassmann manifold of all ν-dimensional linear spaces through the origin in $E_{\nu+N}$. Over $H(\nu, N)$ as base space we can define a bundle of complex vectors by taking the space $S(\nu, N)$ of all pairs (x, E_ν), where E_ν is a ν-dimensional linear subspace of $E_{\nu+N}$ and $x \varepsilon E_\nu$, and defining the projection $P_{\nu N}\colon S(\nu, N) \to H(\nu, N)$ by $P_{\nu N}(x, E_\nu) = E_\nu$. For applications to algebraic geometry it is advantageous to identify all non-zero vectors of $E_{\nu+N}$ which differ from each other by a non-zero complex factor and hence to consider the complex projective space $P_{\nu+N-1}$ of dimension $\nu + N - 1$ and the Grassmann variety $G(\nu - 1, N)$ of all linear spaces of dimension $\nu - 1$ in $P_{\nu+N-1}$. Since $H(\nu, N)$ and $G(\nu - 1, N)$ are homeomorphic in a natural way, this does not make any difference for our present purpose, and we shall use $H(\nu, N)$ as the base space of the universal bundle.

We also remark that the characteristic classes described in the Introduction can be defined for any bundle of complex vector spaces having as base space a finite polyhedron. The bundle described in the last section is called universal, because of the following theorem: [13]

Any bundle of complex vector spaces of dimension ν over a finite polyhedron X is induced by a mapping $f\colon X \to H(\nu, N)$, if $\dim X \leqq 2N$. The bundles induced by two such mappings are equivalent if and only if the mappings are homotopic. If Γ_i, $i = 1, \cdots, \nu$, are the characteristic classes of the universal bundle, then those of the induced bundle are $f^*\Gamma_i$.

On the other hand, the characteristic classes of the universal bundle in $H(\nu, N)$ can be described in a simple way in terms of the homology properties of $H(\nu, N)$. The latter have been studied by Ehresmann,[14] whose results can be summarized as follows: We take a sequence of integers

$$0 \leqq a_1 \leqq a_2 \leqq \cdots \leqq a_\nu \leqq N,$$

and, corresponding to such a sequence, a sequence of linear spaces through the origin 0 in $E_{\nu+N}$:

$$0 \subset L_1 \subset L_2 \subset \cdots \subset L_\nu, \tag{33}$$

such that $\dim L_i = a_i + i$, $i = 1, \cdots, \nu$. A Schubert variety, to be denoted by $(a_1 \cdots a_\nu)$, consists of all the E_ν satisfying the conditions

$$\dim (L_i \cap E_\nu) \geqq i, \qquad i = 1, \cdots, \nu. \tag{34}$$

[13] Chern [1], 88-89, Theorems 1 and 2.

[14] Ehresmann [3].

It is of dimension $\sum_{i=1}^{\nu} a_i$. The Schubert varieties have the following properties, which we shall use and which we state here without proof:

1) Every Schubert variety $(a_1 \cdots a_\nu)$ carries a cycle, whose homology class depends only on the symbol and not on the choice of the sequence (33). All these homology classes form a homology base of $H(\nu, N)$. From this it follows that all odd-dimensional Betti numbers of $H(\nu, N)$ are zero and that there are no torsion coefficients.

2) The intersection number $KI((a_1 \cdots a_\nu), (N - a_\nu, \cdots, N - a_1))$ is equal to one. All other intersection numbers of Schubert varieties of complementary dimensions are zero.

3) Let $(a_1 \cdots a_\nu)$ denote also the cohomology class whose value is one for $(a_1 \cdots a_\nu)$ and is zero for all other homology classes. Then Γ_k is the class $(0 \cdots 0\ 1 \cdots 1)$ consisting of $\nu - k$ zeros followed by k ones.

After these preparations we proceed to the proof of Theorem 4.[15] Let $E_{\nu_1+N_1}$ and $E_{\nu_2+N_2}$ be complex vector spaces and let $H(\nu_1, N_1)$, $H(\nu_2, N_2)$ be the corresponding Grassmann manifolds. Let $f_\alpha\colon X \to H(\nu_\alpha, N_\alpha)$, $\alpha = 1, 2$, be the mappings which induce the bundles B_α over X. We denote by $E_{\nu_1+\nu_2+N_1+N_2}$ the vector space spanned by $E_{\nu_1+N_1}$ and $E_{\nu_2+N_2}$, and by $H(\nu_1 + \nu_2, N_1 + N_2)$ the Grassmann manifold of all linear spaces of dimension $\nu_1 + \nu_2$ through the origin in $E_{\nu_1+\nu_2+N_1+N_2}$. An $E_{\nu_1} \subset E_{\nu_1+N_1}$ and an $E_{\nu_2} \subset E_{\nu_2+N_2}$ span an $E_{\nu_1+\nu_2} \subset E_{\nu_1+\nu_2+N_1+N_2}$. This defines a mapping

$$F\colon H(\nu_1, N_1) \times H(\nu_2, N_2) \to H(\nu_1 + \nu_2, N_1 + N_2). \tag{35}$$

Then the bundle $B_1 \boxtimes B_2$ over X is induced by the mapping $F \circ f$, where

$$f\colon X \to H(\nu_1, N_1) \times H(\nu_2, N_2) \tag{36}$$

is defined by $f(x) = (f_1(x), f_2(x))$, $x \in X$. Denote by $\Gamma_1^{(1)}, \cdots, \Gamma_{\nu_1}^{(1)}$; $\Gamma_1^{(2)}, \cdots, \Gamma_{\nu_2}^{(2)}$; $\Gamma_1, \cdots, \Gamma_{\nu_1+\nu_2}$ the characteristic classes of the universal bundles over $H(\nu_1, N_1)$, $H(\nu_2, N_2)$, $H(\nu_1 + \nu_2, N_1 + N_2)$ respectively. We assert that the duality theorem is a consequence of the formula

$$F^*\Gamma_k = \sum_{\substack{0 \leqq i \leqq k \\ k-\nu_2 \leqq i \leqq \nu_1}} \Gamma_i^{(1)} \otimes \Gamma_{k-i}^{(2)}, \qquad k = 1, \cdots, \nu_1 + \nu_2. \tag{37}$$

In fact, we have

$$C_k = f^*F^*\Gamma_k, \qquad C_i^{(1)} = f_1^*\Gamma_i^{(1)}, \qquad C_j^{(2)} = f_2^*\Gamma_j^{(2)},$$

$$C_i^{(1)} \cup C_j^{(2)} = (f_1^*\Gamma_i^{(1)}) \cup (f_2^*\Gamma_j^{(2)}) = f^*(\Gamma_i^{(1)} \otimes \Gamma_j^{(2)}).$$

---

[15] The idea of this proof was indicated by Wu: cf. Wu [17].

Applying the dual homomorphism f^* to (37), we get

$$C_k = \sum_{\substack{0 \leqq i \leqq k \\ k-\nu_2 \leqq i \leqq \nu_1}} C_i^{(1)} \cup C_{k-i}^{(2)}, \tag{38}$$

which is obviously an equivalent formulation of the duality theorem.

To establish (37) it suffices to prove that its two sides, being cohomology classes of $H(\nu_1, N_1) \times H(\nu_2, N_2)$, have the same value for a homology base of dimension k of $H(\nu_1, N_1) \times H(\nu_2, N_2)$. Such a base is provided by the products of Schubert varieties of $H(\nu_1, N_1)$ and $H(\nu_2, N_2)$, whose dimensions have the sum k. Consider first a product $\zeta_1 \times \zeta_2$, of which at least one factor is not of the form $(0 \cdots 0\ 1 \cdots 1)$. The value of the right-hand side of (37) for $\zeta_1 \times \zeta_2$ is then zero, while we have

$$(F^*\Gamma_k) \cdot (\zeta_1 \times \zeta_2) = \Gamma_k \cdot F(\zeta_1 \times \zeta_2) = KI(\eta_k, F(\zeta_1 \times \zeta_2)),$$

where

$$\eta_k = (\underbrace{N_1 + N_2 - 1 \cdots N_1 + N_2 - 1}_{k}\ N_1 + N_2 \cdots N_1 + N_2). \tag{39}$$

To prove that this intersection number is zero, it suffices to assume the two Schubert varieties to be in general position and show that they have no element in common. In fact, from our assumption on $\zeta_1 \times \zeta_2$, each of the elements of $F(\zeta_1 \times \zeta_2)$ passes through a fixed linear space A of dimension $\nu_1 + \nu_2 - k + 1$. On the other hand, a linear space $E_{\nu_1+\nu_2}$ belongs to η_k if and only if it has a linear space of dimension $\geqq k$ in common with a fixed linear space B of dimension $N_1 + N_2 + k - 1$. We can choose B so that A and B have only the zero vector in common. If there is an $E_{\nu_1+\nu_2}$ through A which has a linear space of dimension $\geqq k$ in common with B it would follow that $A \cap B$ is of dimension $\geqq 1$. It follows therefore that the intersection number in question is zero.

It remain to take $\zeta_1 = (0 \cdots 1\ \underbrace{1 \cdots 1}_{i})$, $\zeta_2 = (0 \cdots 0\ \underbrace{1 \cdots 1}_{k-i})$, for a fixed i, and to prove that, under this choice,

$$KI(\eta_k, F(\zeta_1 \times \zeta_2)) = 1.$$

Let $L_{\nu_1-i}, L_{\nu_1+1}, M_{\nu_2-k+i}, M_{\nu_2+1}$ be linear spaces, with subscripts indicating their dimensions, which are used to define ζ_1, ζ_2, so that

$$L_{\nu_1-i} \subset L_{\nu_1+1} \subset E_{\nu_1+N_1},$$

$$M_{\nu_2-k+i} \subset M_{\nu_2+1} \subset E_{\nu_2+N_2}.$$

By the condition (34) for Schubert varieties the elements $V_{\nu_1} \varepsilon \zeta_1$, $V_{\nu_2} \varepsilon \zeta_2$ are respectively characterized by

$$L_{\nu_1-i} \subset V_{\nu_1} \subset L_{\nu_1+1},$$

$$M_{\nu_2-k+i} \subset V_{\nu_2} \subset M_{\nu_2+1}.$$

An element $V_{\nu_1+\nu_2} \varepsilon F(\zeta_1 \times \zeta_2)$ is spanned by V_{ν_1} and V_{ν_2}. The condition that it belongs also to η_k is

$$\dim (V_{\nu_1+\nu_2} \cap B) \geqq k.$$

Since the L's, M's, and B are supposed to be in general position, B has only the zero vector in common with the space $C_{\nu_1+\nu_2-k}$ of dimension $\nu_1 + \nu_2 - k$ spanned by L_{ν_1-i} and M_{ν_2-k+i}. There exists therefore a space $D_{N_1+N_2+k}$ of dimension $N_1 + N_2 + k$, which contains B and is supplementary to $C_{\nu_1+\nu_2-k}$ in $E_{\nu_1+\nu_2+N_1+N_2}$. The projections of L_{ν_1+1} and M_{ν_2+1} in $D_{N_1+N_2+k}$ from $C_{\nu_1+\nu_2-k}$ are of dimensions $i+1$ and $k-i+1$ respectively, which we denote by L_{i+1}' and M_{k-i+1}'. Their intersections with B, say L_i'' and M_{k-i}'', are then of dimensions i and $k-i$ respectively. For a $V_{\nu_1+\nu_2}$ spanned by $V_{\nu_1} \varepsilon \zeta_1$, $V_{\nu_2} \varepsilon \zeta_2$ to belong to η_k, or, what is the same, to have an intersection of dimension $\geqq k$ with B, it is therefore necessary and sufficient that it contains L_i'' and M_{k-i}''. Such a $V_{\nu_1+\nu_2}$ is uniquely determined, as the space spanned by L_{ν_1-i}, M_{ν_2-k+i}, L_i'', M_{k-i}''. This proves that $F(\zeta_1 \times \zeta_2)$ and η_k have in common exactly one $V_{\nu_1+\nu_2}$.

Using a coordinate system, we can study, in the differentiable manifold $H(\nu_1 + \nu_2, N_1 + N_2)$, the submanifolds η_k and $F(\zeta_1 \times \zeta_2)$ in the neighborhood of their intersection. It is easily seen that the tangent spaces of these submanifolds are skew to each other, so that the points of intersection is to be counted simply. This completes the proof of formula (37) and hence the duality theorem.

6. Proofs of Theorems 5 and 6. Theorem 5 is an easy consequence of the duality theorem. To prove it we put an Hermitian metric on M, so that the structural group of the tangent bundle reduces to the unitary group. To a point $y \varepsilon B_{n1}$, that is, a tangent direction of M, let $T(y)$ be the subspace of dimension $n-1$ of the tangent vector space which is perpendicular to y. The vectors of $T(y)$, for all $y \varepsilon B_{n1}$, form a bundle of complex vector spaces of dimension $n-1$ over B_{n1}. The product of this bundle and the bundle $p_{n1}\colon \bar{B}_{n1} \to B_{n1}$ is clearly equivalent to the bundle over B_{n1} induced by the mapping $q_{n1}\colon B_{n1} \to M$. If we denote by

$$C(t) = \sum_{i=0}^{n} C_i t^i, \qquad C_0 = 1,$$

the characteristic polynomial in M, the characteristic polynomial of the induced bundle is

$$q_{n1}{}^*C(t) = \sum_{i=0}^{n} q_{n1}{}^*(C_i)t^i.$$

On the other hand, the characteristic polynomial of the bundle $p_{n1}\colon \tilde{B}_{n1} \to B_{n1}$ is $1 + u_1 t$, while that of the other factor is of the form

$$D(t) = \sum_{j=0}^{n-1} D_j t^j, \qquad D_0 = 1.$$

By the duality theorem we have therefore, identically in t,

$$\sum_{i=0}^{n} q_{n1}{}^*(C_i)t^i = (\sum_{i=0}^{n-1} D_i t^i)(1 + u_1 t).$$

Equating the corresponding coefficients of t^i, we get

$$q_{n1}{}^*(C_i) = D_i + u_1 D_{i-1}, \qquad i = 1, \cdots, n,$$

where we define $D_{-1} = D_n = 0$. Elimination of the D's gives formula (10). This proves Theorem 5.

In order to prove Theorem 6, we need the following algebraic lemma:

LEMMA 6.1. *Let*

$$C(t) = 1 + C_1 t + \cdots + C_n t^n \tag{40}$$

be a polynomial in t with coefficients in a commutative ring with unit element 1. To the elements C_i introduce $\bar{C}_k$, $k = 1, 2, \cdots$ as the coefficients of the formal power series

$$\bar{C}(t) = \sum_{k=0}^{\infty} \bar{C}_k t^k, \qquad \bar{C}_0 = 1, \tag{41}$$

so that

$$C(t)\bar{C}(t) = 1. \tag{42}$$

Let u be an element of the ring satisfying the equation

$$u^n C(-\frac{1}{u}) \equiv u^n(1 - \frac{C_1}{u} + \cdots + (-1)^n \frac{C_n}{u^n}) = 0. \tag{43}$$

Then we have, for any integer $N \geqq 1$,

$$u^{N+n-1} = (-1)^N \bar{C}_N u^{n-1} + \text{lower powers of } u. \tag{44}$$

To prove the lemma we write

$$\begin{aligned} 1 - C(t)(1 + \bar{C}_1 t + \cdots + \bar{C}_{N-1}t^{N-1}) &= C(t)(\bar{C}_N t^N + \cdots) \\ &= t^N D(t) + t^{n+N} P(t), \end{aligned}$$

where $P(t)$ is a power series in t and $D(t)$ a polynomial in t of degree $\leqq n-1$:

$$D(t) = D_0 + D_1 t + \cdots + D_{n-1} t^{n-1}.$$

Comparing the coefficients of t^N in both sides, we get

$$\bar{C}_N = D_0.$$

We now put $t = -1/u$ in the last identity and multiply the two sides by u^{n+N-1}. This gives

$$u^{n+N-1} - \{u^n C(-\frac{1}{u})\}\{u^{N-1} - \bar{C}_1 u^{N-2} + \cdots + (-1)^{N-1}\bar{C}_{N-1}\}$$
$$= (-1)^N u^{n-1} D(-\frac{1}{u}) + (-1)^{n+N}\frac{1}{u}P(-\frac{1}{u}).$$

This is a formal identity in the positive and negative powers of u. Restricting ourselves to the non-negative powers and taking account of the hypothesis of the Lemma, we get

$$u^{n+N-1} = (-1)^N u^{n-1} D(-\frac{1}{u}).$$

From this the Lemma follows.

We proceed now to the proof of Theorem 6. As in the proof of Theorem 5, we consider the bundle induced over B_{nr} by the mapping $q_{nr}\colon B_{nr} \to M$. Since M has an Hermitian metric, an element $L_0 \subset L_1 \subset \cdots \subset L_{r-1}$ of B_{nr} determines, and can be determined by, r mutually perpendicular directions $y_1, \cdots, y_r$ with the same origin, such that $y_1, \cdots, y_i$ span L_{i-1}, $i = 1, \cdots, r$. The induced bundle is therefore equivalent to the product of a bundle of vector spaces of dimension $n-r$ and the bundles of one-dimensional vector spaces over each of $y_1, \cdots, y_r$. We denote by

(45) $$k(t) = k_0 + k_1 t + \cdots + k_{n-r} t^{n-r}, \qquad k_0 = 1,$$

and $1 + u_i t$, $i = 1, \cdots, r$, their respective characteristic polynomials. From the duality theorem we have

(46) $$D(t) \equiv \sum_{i=0}^{n} D_i t^i = k(t) \prod_{j=1}^{r} (1 + u_j t),$$

where

(47) $$D_i = q_{nr}^*(C_i), \qquad i = 0, \cdots, n.$$

It is easy to see that the classes u_i are related to the classes v_i by the relations

(14) $$v_i = u_1 + \cdots + u_i, \qquad i = 1, \cdots, r,$$

as given in the Introduction. We proceed to derive from (46) a formula for the expression under the parenthesis in the left-hand side of (15). In what follows, multiplication of cohomology classes will be understood in the sense of cup product; it is commutative, because the classes concerned are all even-dimensional.

We introduce the polynomial

$$D'(t) = \sum_{i=0}^{n-r+1} D_i' t^i = (1 + u_r t) k(t), \tag{48}$$

and define the formal power series

$$\begin{aligned} \bar{D}'(t) &= \sum_{i=0}^{\infty} \bar{D}_i' t^i, & \bar{D}_0' &= 1, \\ \bar{D}(t) &= \sum_{i=0}^{\infty} \bar{D}_i t^i, & \bar{D}_0 &= 1, \end{aligned} \tag{49}$$

by means of the relations

$$D(t)\bar{D}(t) = 1, \qquad D'(t)\bar{D}'(t) = 1. \tag{50}$$

Then we have

$$D(t)\bar{D}'(t) = (1 + u_1 t) \cdots (1 + u_{r-1} t).$$

This equation completely determines $\bar{D}'(t)$. It follows by observation that

$$\bar{D}'(t) = \bar{D}(t)(1 + u_1 t) \cdots (1 + u_{r-1} t).$$

From the definition of $D'(t)$ we get

$$u_r^{n-r+1} D'(-\frac{1}{u_r}) = 0.$$

By Lemma 6. 1, we have therefore

$$u_r^{2n-r} = (-1)^n \bar{D}'_n u_r^{n-r} + \text{terms in lower powers of } u_r.$$

But

$$\bar{D}'_n = u_1 \cdots u_{r-1} \bar{D}_{n-r+1} + \cdots,$$

where the unwritten terms are of degrees $< r - 1$ in $u_1, \cdots, u_{r-1}$. It follows that

$$u_1^{n-2} \cdots u_{r-1}^{n-r} u_r^{2n-r} = (-1)^n \bar{D}_{n-r+1} u_1^{n-1} \cdots u_{r-1}^{n-r+1} u_r^{n-r} + \cdots.$$

Under the homomorphism Ψ_{nr} the unwritten terms go to 0, while the first term goes to $(-1)^n \bar{C}_{n-r+1}$. This proves Theorem 6.

7. Application to algebraic geometry. We now sketch briefly the application of the above results to the case in which M is a non-singular algebraic variety of dimension n in a complex projective space of higher dimension. For this purpose we introduce another bundle over M. It is the bundle $[\tilde{B}_{nr}]$ obtained from $\tilde{B}_{nr}$ by adjoining to each fiber of $\tilde{B}_{nr}$, which is a Cartesian product of r complex lines with the origins deleted, the origin and a point at infinity. Its fiber is therefore a Cartesian product of r complex projective lines $L_1 \times \cdots \times L_r$. It is acted on, in an intransitive manner, by the general linear group $G(n)$, so that $[\tilde{B}_{nr}]$ is also an associated bundle of the principal bundle over M with $G(n)$ as structural group.

At this point we have to distinguish between the notion of an algebraic variety in the classical sense as one imbedded in the projective complex space and that of an abstract variety in the sense of Weil [16] defined by means of "overlapping neighborhoods." The variety $B_{n,n}^*$, being the principal bundle, is clearly an abstract variety. By the theory of fiber bundles in algebraic geometry, which is entirely analogous to the topological case, it follows that the bundles B_{nr}, $[\tilde{B}_{nr}]$, as associated bundles, are also abstract varieties and that the projections $q_{nr}\colon B_{nr} \to M$, $p_{nr}\colon \tilde{B}_{nr} \to B_{nr}$ are rational mappings.

On $[\tilde{B}_{nr}]$ there are sub-varieties V_{i0}, $V_{i\infty}$, $i = 1, \cdots, r$, defined by setting to zero or to ∞ the i-th coordinate in the fiber. Utilizing these sub-varieties, we can define the classes v_i in B_{nr} according to the following lemma, which was first given by Weil: [17]

LEMMA 7.1. *To an analytic cross-section f of the bundle $p_{nr}\colon \tilde{B}_{nr} \to B_{nr}$ the cycle*

$$(p_{nr})_*(f(B_{nr}) \cdot V_{i\infty} - f(B_{nr}) \cdot V_{i0})$$

is defined up to linear equivalence. Its homology class is $\mathfrak{D} v_i$.

To prove Theorem 7 we need also the following lemma:

LEMMA 7.2. *On a non-singular algebraic variety the intersection class of a finite number of divisor classes contains an algebraic cycle.*

For classical algebraic varieties this follows by induction from the following statement: If Z is an algebraic cycle on M and D is a divisor in M, there exists a divisor D' which is equivalent to D and which intersects Z properly.

[16] Weil [15], Chapter VII.

[17] Weil [16].

In fact, let F be a hypersurface in the ambient projective space of M, whose intersection with M consists of D and a divisors D_1 which does not contain a given point of Z.[18] Then D_1 intersetcs Z properly. Let F_1 be a hypersurface, of the same degree as F, which intersects both M and Z properly. Set $D_2 = F_1 \cdot M$. Then the divisor $D' = D_2 - D_1$ is equivalent to D and intersects Z properly.

From these two lemmas the proof of Theorem 7 follows immediately. For $\mathfrak{D} v_1, \cdots, \mathfrak{D} v_r$, and hence $\mathfrak{D} u_1, \cdots, \mathfrak{D} u_r$, contain divisor classes. Therefore $(\mathfrak{D} u_1)^{n-2} \cdots (\mathfrak{D} u_{r-1})^{n-r} (\mathfrak{D} u_r)^{2n-r}$ contains an algebraic cycle. Its projection in M, which is a cycle belonging to the homology class $(-1)^n \bar{C}_{n-r+1}$, is an algebraic cycle. This proves Theorem 7.

To make our proof complete, one should prove Lemma 7.2 for abstract varieties. While this is "undoubtedly" true, no proof of it has been published. Alternately, it would presumably be easy to imbed the fiber bundles B_{nr}, $[\tilde{B}_{nr}]$ into a projective space, and then the theorem would again follow. These are questions of algebraic geometry which cannot be completely discussed without a lengthy introduction. We shall therefore not take them up in this paper.

BIBLIOGRAPHY.

[1] S. S. Chern, "Characteristic classes of Hermitian manifolds," *Annals of Mathematics*, (2), vol. 47 (1946), pp. 85-121.

[2] M. Eger, "Sur les systèmes canoniques d'une variété algébrique à plusieurs dimensions," *Annales Scientifique de l'École Normale Superieure*, (3), vol. 60 (1943), pp. 143-172.

[3] C. Ehresmann, "Sur la topologie de certains espaces homogènes," *Annals of Mathematics*, (2), vol. 35 (1934), pp. 396-443.

[4] S. Eilenberg and S. MacLane, "Group extensions and homology," *ibid.*, (2), vol. 43 (1942), pp. 757-831.

[5] ——— and N. Steenrod, *Foundations of algebraic topology*, Princeton (1952).

[6] W. Gysin, "Zur Homologietheorie der Abbildungen und Faserungen der Mannigfaltigkeiten," *Commentarii Mathematici Helvetici*, vol. 14 (1941), pp. 61-122.

[7] G. Hirsch, "L'anneau de cohomologie d'un espace fibré et les classes caractéristiques," *Comptes rendus de l'Académie des Sciences*, vol. 229 (1949), pp. 1297-1299.

[18] van der Waerden [14], 635.

[8] ———, "Quelques relations entre l'homologie dans les espaces fibrés et les classes caractéristiques relatives à un groupe de structure," *Colloque de Topologie*, Paris (1951), pp. 123-136.

[9] W. V. D. Hodge, "The characteristic classes of algebraic varieties," *Proceedings of the London Mathematical Society*, (3), vol. 1 (1951), pp. 138-151.

[10] J. Leray, "L'homologie d'un espace fibré dont la fibre est connexe," *Journal de Mathematiques*, vol. 29 (1950), pp. 169-213.

[11] E. H. Spanier, "Homology theory of fiber bundles," *Proceedings of the International Congress of Mathematicians*, vol. II (1950), pp. 390-396.

[12] N. Steenrod, *Topology of fibre bundles*, Princeton (1951).

[13] J. A. Todd, "The geometrical invariants of algebraic loci," *Proceedings of the London Mathematical Society*, (2), vol. 43 (1937), pp. 127-138.

[14] B. L. van der Waerden, "Zur algebraischen Geometrie XIV, Schnittpunktszahlen von algebraischen Mannigfaltigkeiten," *Mathematische Annalen*, vol. 115 (1938), pp. 619-642.,

[15] A. Weil, *Foundations of algebraic geometry*, New York (1946).

[16] ———, "Fibre-spaces in algebraic geometry," *Conference on algebraic geometry and algebraic number theory*, Chicago (1949), pp. 55-59.

[17] W. T. Wu, "Les *i*-carrés dans une variété grassmannienne," *Comptes rendus de l'Académie des Sciences*, Paris, vol. 230 (1950), pp. 918-920.

ON THE INVERSE OF THE PARABOLIC DIFFERENTIAL OPERATOR $\partial^2/\partial x^2 - \partial/\partial t$.*

By Philip Hartman and Aurel Wintner.

1. Let R denote a (sufficiently small) simply connected region in an (x, y)-plane, say the rectangle $R = R(1)$, where

$$(1) \qquad R(c): \; (-1 \leqq x \leqq 1, 0 \leqq y \leqq c),$$

and let R^0 be the interior of $R(1)$,

$$(2) \qquad R^0: \; (-1 < x < 1, 0 < y < 1).$$

If $u = u(x, y)$ is a (real-valued) function on R^0, put

$$(3) \qquad \partial u = u_{xx} - u_y; \qquad\qquad (3\text{ bis}) \qquad \Delta u = u_{xx} + u_{yy},$$

if the partial derivatives occurring in the respective operators ∂, Δ exist for $u = u(x, y)$. In contrast, denote by $\partial^* u = \partial^* u(x, y)$, $\Delta^* u = \Delta^* u(x, y)$ the operators

$$(4) \qquad \partial^* u = \lim \partial_h u; \qquad\qquad (4\text{ bis}) \qquad \Delta^* u = \lim \Delta_h u,$$

if the respective limits (4), (4 bis), where $h \rightarrow 0$, exist for the quotients $\partial_h u$, $\Delta_h u$ defined by

$$(5) \qquad \begin{aligned} h^2 \cdot \partial_h u &= u(x+h, y) + u(x-h, y) + u(x, y-h^2) - 3u(x, y) \\ &\equiv u(x+h, y) - 2u(x, y) + u(x-h, y) \\ &\qquad - \{u(x, y) - u(x, y-h^2)\} \end{aligned}$$

and

$$(5\text{ bis}) \qquad \begin{aligned} h^2 \Delta_h u = u(x+h, y) &+ u(x, y+h) + u(x-h, y) \\ &+ u(x, y-h) - 4u(x, y). \end{aligned}$$

The expression (4 bis)-(5 bis), formed in analogy to the contracted second derivative (Riemann, Hölder) of a function of a single variable, was introduced by Zaremba [9] as an extension of (3 bis), whereas Gevrey's corresponding extension of (3) is not (4)-(5) but a limit which presupposes the existence of u_x; cf. [1], p. 363.

* Received February 26, 1953.

The definition (4) of $\partial^* u$ is unsymmetric. A more natural (but apparently more restrictive) definition results if (5) is replaced by

$$\partial_h(u) = \{u(x+h, y) - 2u(x, y) + u(x-h, y)\}/h^2$$
$$- \{u(x, y) - u(x, y-k)\}/k,$$

where $|k| = h^2 > 0$. It turns out that the limit (4) exists and is continuous when $\partial_h u$ is defined by (5) if and only if the same holds when $\partial_h u$ is defined in the more symmetric manner; cf. the proof of (i*) and (ii*) below.

Let $G = G(x, y; x', y')$ denote the function $\gamma(s, t)$ of

$$(6) \qquad s = x - x' \quad \text{and} \quad t = y - y'$$

defined for $(s, t) \neq (0, 0)$ by

$$(7) \qquad \gamma \equiv 0 \text{ if } t \leqq 0 \text{ and } \gamma = t^{-\frac{1}{2}} \exp(-s^2/4t) \text{ if } t > 0;$$

so that

$$(7 \text{ bis}) \qquad G(x, y; x', y') = \gamma(x - x', y - y').$$

It is classical that (7) annihilates, for fixed (x, y) distinct from (x', y'), the adjoint $\partial^2/\partial x'^2 + \partial/\partial y'$ of $\partial = \partial^2/\partial x^2 - \partial/\partial y$ and that, correspondingly,

$$(8) \qquad u(x, y) = -\tfrac{1}{2}\pi^{-\frac{1}{2}} \iint_R G(x, y; x', y') f(x', y')\, dx' dy'$$

represents *a* solution of $\partial u = f(x, y)$ on the interior (2) of (1) if $f(x, y)$ is a sufficiently smooth function on R (the factor $-\frac{1}{2}\pi^{-\frac{1}{2}}$ in (8) corresponds to the "infinity" of the "source function" γ). Thus (7 bis) and the logarithm of the distance between (x, y) and (x', y') play analogous parts for the respective inhomogeneous extensions,

$$(9) \qquad \partial u = f(x, y); \qquad\qquad (9 \text{ bis}) \qquad \Delta u = f(x, y),$$

of the equations of Fourier and of Laplace,

$$(10) \qquad \partial u = 0; \qquad\qquad (10 \text{ bis}) \qquad \Delta u = 0.$$

It is well-known that *if $f(x, y)$ is just continuous* on R (instead of satisfying something like a Hölder condition), then (9 bis) cannot be handled directly but must be replaced by the second of the equations

$$(11) \qquad \partial^* u = f(x, y); \qquad\qquad (11 \text{ bis}) \qquad \Delta^* u = f(x, y),$$

to which end the theory of (10 bis) must be extended to the second of the equations

$$(12) \qquad \partial^* u = 0; \qquad\qquad (12 \text{ bis}) \qquad \Delta^* u = 0.$$

It appears however that, while the relevant local theory of (9 bis)-(12 bis), as well as that of the second of the homogeneous equations

$$(13)\qquad \partial u + fu = 0; \qquad\qquad (13\text{ bis})\qquad \Delta u + fu = 0,$$

is fully developed (Zaremba [9], Petrini [6]; cf. Wintner [8]), the results available for (9)-(13) in the literature are far from being of comparable sharpness. The Hölder criterion of E. E. Levi [5], p. 239, and its generalization by Gevrey [1], pp. 350-352, supply only sufficient (and, as will be seen below, surely not necessary) conditions for the existence of the relevant derivatives of the formal solution (8) of (9), in contrast to Petrini's complete results on the corresponding questions on the logarithmic potential of a continuous density $f(x, y)$. The object of this paper is to fill out these gaps in the theory of (9)-(13).

2. By a solution $u = u(x, y)$ of (10) on (2) will be meant any function for which the derivatives u_{xx}, u_y, those occurring in (3), exist and are equal at every point of (2). Correspondingly, by a solution of (12) on (2) will be meant a function $u(x, y)$ defined on (10) in such a way that the limit (4) of (5) exists at every point of (2) and vanishes identically. The "solutions" of (11), (12) or (13), where $f = f(x, y)$ is any given (usually continuous) function, are defined similarly.

If u is a solution of $\partial u = 0$, then u_y and u_x exist but u need not be continuous. In fact, there exist on R^0 solutions u of $\partial u = 0$ which are discontinuous at every point of a dense subset (even of the second category) of R^0; cf. [8], p. 732 and the corresponding function ϕ in [2], p. 371, top. On the other hand, if u is of class C^1 on R^0, then, while all that it is clear from $u_{xx} = u_y$ is that u_{xx} is continuous, it turns out (cf. [4], p. 2, where reference is given to an earlier paper of Holmgren) that u must be analytic in x for fixed y, and of class C^∞ (though not in general analytic) in x and y together (hence in y for fixed x). Actually, the same is true if nothing beyond the continuity (or the local boundedness) of u is assumed:

(i) *If $u(x, y)$ is a solution of $\partial u = 0$ on R^0 and if $u(x, y)$ is locally bounded and measurable (for example, if u is continuous on R^0), then u is of class C^1 on R^0 (hence of class C^∞ and, for fixed y, analytic in x).*

It was pointed out in [2], pp. 368-369, that, under the assumption that u is continuous on R^0, assertion (i) is between the lines of the classical literature. A verification of (i) under this condition was sketched in [2], pp. 368-369. This proof depended on a maximum principle. The proof

(although not the statement) of this principle was given by Gevrey [1], pp. 372-374. Actually, the proof of the required maximum principle does not depend on the continuity of u, but only on the continuity of u on each of the lines $x = \text{const.}$ and $y = \text{Const.}$ (and the local boundedness of u). Hence (i) follows by an obvious modification of the proof of (I) in [2].

The assertion (i) has the following extension:

(i*) *If $u(x, y)$ is a continuous solution of $\partial^* u = 0$ on R^0, then it is of class C^1 (hence, a solution of $\partial u = 0$) on R^0.*

The assumption of continuity on u can be replaced by assumptions of local boundedness and measurability.

Clearly, (i*) leads, via (i), to the following corollary:

(i* bis) *If $u = v(x, y)$ and $u = w(x, y)$ are two continuous solutions of $\partial^* u = f(x, y)$, then the function $u = v - w$ is of class C^∞ (which means that $v - w$ is a solution of $\partial u = 0$, rather than just of $\partial^* u = 0$).*

3. In view of (3) and (4)-(5), this corollary of (i*) implies an enumeration of all continuous solutions u of $\partial u = f$ in terms of one of them, say of $u = u_0$, *if* there exists such a u_0. But this assumption need not be satisfied, not even if f is continuous:

(ii) *There exist on R^0 continuous functions $f(x, y)$ corresponding to which $\partial u = f$ has no continuous solution $u(x, y)$ on any open subset of R^0.*

The same holds when (9) is replaced by (13):

(ii′) *The assertion of* (ii) *on the inhomogeneous equation $\partial u = f(x, y)$ holds for the homogeneous equation $\partial u + f(x, y) u = 0$ also, if the solution $u \equiv 0$ of the latter is disregarded.*

This has the following variant:

(ii″) *There exist on R positive, continuous functions f corresponding to which the differential equation $u_{xx} - f(x, y) u_y = 0$ fails to possess any continuous solution $u(x, y) \not\equiv \text{const.}$ on any subdomain of R.*

The point in the transition from the differential operator ∂ to its "Abelian" form ∂^* is not revealed by the above theorems. It is made clear however if (ii) is contrasted with the following existence theorem:

(ii*) *If $f(x, y)$ is continuous on the closure R of R^0, then $\partial^* u = f$ possesses on R^0 continuous solutions $u(x, y)$; in fact,* (8) *is such a solution* (whence *all* continuous solutions follow by (i* bis)).

4. The following theorems indicate the boundary between $\partial u = f$ and $\partial^* u = f$ more sharply than do (ii) and (ii*):

(iii) *If f is continuous on R, and if u is the particular solution* (8) *of $\partial^* u = f$ on R^0, then*

(iiiα) *u is uniformly continuous on R^0 (that is, continuous on R);*

(iiiβ) *$u_y = \partial u/\partial y$ need not exist; although*

(iiiγ) *$D^{1-\epsilon}{}_y u = \partial^{1-\epsilon} u/\partial y^{1-\epsilon}$* (Liouville) *exists and is continuous on R^0 for every $\epsilon > 0$;*

(iiia) *$u_x = \partial u/\partial x$ exists and is continuous on R^0;*

(iiib) *$u_{xx} = \partial^2 u/\partial x^2$ need not exist; although*

(iiic) *$D^{1-\epsilon}{}_x u_x = \partial^{2-\epsilon} u/\partial u^{2-\epsilon}$ exists and is continuous on R_0 if $\epsilon > 0$.*

According to (iiia) and (iiiγ), both operators

$$\partial/\partial x \pm \partial^{\frac{1}{2}}/\partial y^{\frac{1}{2}} \tag{14}$$

are applicable to every u, which is of formal interest, since ∂ is the product of the two (formally commutable) operators (14); cf. (3).

That the function (8) has the properties (iiiα) and (iiia) is contained in E. E. Levi [5], pp. 229-235; cf. Gevrey [1], pp. 343-344. That (8) satisfies (iiiγ) and (iiic) follows from considerations of Gevrey [1], pp. 357-360, in view of the connection (Weyl [7]) between Hölder conditions and fractional differentiation. Hence only the equivalent parts (iiiβ), (iiib) of (iii) have to be proved.

In connection with (iiiβ) and (iiib), it is worth mentioning that, *at a given point (x, y), either both of the derivatives u_y, u_{xx} exist or neither of them does. Furthermore, when they exist at a given point, then $\partial u = f$.* This is implied by a result of Gevrey [1], p. 369; cf. (iv) below and its proof.

The analogue for (11) of Petrini's theorem for (11 bis) is as follows:

(iv) *Let $f(x, y)$ be continuous on the closure of R^0 and let $u(x, y)$ be a continuous solution of $\partial^* u = f$ on R^0. Then necessary and sufficient for the existence of $u_y(x, y)$ and/or $u_{xx}(x, y)$ at a point (x, y) of R^0 is the existence of*

$$\lim_{\epsilon \to 0} \iint_D f(x', y') G_y(x, y; x', y') \, dx' dy', \tag{15}$$

where $D = D(\epsilon)$ denotes either the set of points (x', y') of R for which

$y' < y - \epsilon$ or the subset of R outside the rectangle $|x' - x| < \epsilon^{\frac{1}{2}}$, $y > y' > y - \epsilon$. If the limit (15) *exists, then $\partial u = f$ holds at* (x, y).

Since the limit (15) exists when f is a constant, it follows that (15) exists if and only if

(15 bis) $L = \lim_{\epsilon \to 0} L_\epsilon$, where

$$L_\epsilon = -\tfrac{1}{2}\pi^{-\frac{1}{2}} \iint_D \{f(x', y') - f(x, y)\} G_y(x, y; x', y') dx'dy',$$

does; in which case, at the point (x, y),

$$u_y = z_y f + L \quad \text{and} \quad u_{xx} = f + z_y f + L, \tag{16}$$

where $z = z(x, y)$ is the function (8) belonging to $f \equiv 1$.

It will be noted that, corresponding to Petrini's criteria for (3 bis), the necessary and sufficient criterion (15) supplied by (iv) deals with the question of existence, rather than with the continuity, of the partial derivatives occurring in the differential operator. It will however be clear from the proof of (iv) that, in order to assure the existence of continuous derivatives u_{xx}, u_y on (2), it is sufficient to require the satisfaction of (15) uniformly on every compact subset of (2). This leads to the following:

(iv bis) *In order that some and/or every continuous solution $u = u(x, y)$ of $\partial^* u = f$ on R^0 have continuous partial derivatives u_{xx}, u_y, it is sufficient that there exists a* (*continuous*) *function $\psi(s, t)$ defined for small $s \geqq 0$, $t \geqq 0$, satisfying*

$$\int_{+0} \int_{+0} \psi(s, t) \, | \gamma_t(s, t) | \, ds dt < \infty$$

and

$$| f(x, y) - f(x', y') | \leqq \psi(| x - x' |, | y - y' |).$$

5. *Proof of* (i*). The definition (4) of $\partial^* u$ and the first representation of $h^2 \partial_h u$ in (5) show that if $\partial^* u$ exists and is positive on some square S: $| x - x_0 | \leqq \epsilon$, $| y - y_0 | \leqq \epsilon$ contained in R^0, then the maximum of u on S is assumed on the lateral $(x = x_0 \pm \epsilon)$ or the lower $(y = y_0 - \epsilon)$ boundaries of S. Since u is continuous, there exists a (unique) continuous function $v = v(x, y)$ on S such that v (is of class C^∞ and) satisfies $\partial v = 0$ in the interior of S and $v \equiv u$ on the lateral and lower boundaries of S; cf., e. g., (IV) in [2], p. 369. There also exists a function $w = w(x, y)$ of

class C^∞ on S satisfying $\partial w = 1$, while $w = 0$ on the lateral and lower boundaries of S.

The proof of (i*) can now be completed by the arguments used by Zaremba [9], pp. 146-147, to pass from $\Delta^* u = 0$ to $\Delta u = 0$. Thus, if $\lambda > 0$ is an arbitrary constant, then either of the two functions $U = \pm (u - v) + \lambda w$ is continuous on S, vanishes on the lateral and lower boundaries of S and satisfies $\partial^* U = \lambda > 0$ on the interior of S. By the above maximum principle for ∂^*, it follows that $U \leqq 0$, that is, $| u - v | \leqq \lambda w$ on S. On letting $\lambda \to 0$, it is seen that $u \equiv v$ on S.

6. *Preliminary estimates.* In what follows, the abbreviation $G(x, y)$ will be used for $G(x, y; x', y')$, and ϵ will denote a (small) positive number.

LEMMA. *Let $I_k(\epsilon) = I_k(\epsilon; x, y)$, where $k = 1, 2, 3, 4$, denote the respective integrals*

$$(17_1) \qquad I_1(\epsilon) = \int_{y-\epsilon}^{y} \int_{-1}^{1} G(x, y) dx' dy',$$

$$(17_2) \qquad I_2(\epsilon) = \int_{0}^{y-\epsilon} \int_{-1}^{1} | G_{yy}(x, y) | \, dx' dy',$$

$$(17_3) \qquad I_3(\epsilon) = \int_{0}^{y-\epsilon} \int_{-1}^{1} | G_{xxx}(x, y) | \, dx' dy',$$

$$(17_4) \qquad I_4(\epsilon) = \int_{y-\epsilon}^{y} \int_{|x'-x| \geqq \delta} | G_y(x, y) | \, dx' dy', \qquad \text{where } \delta = \epsilon^{\frac{1}{3}} > 0.$$

Then, uniformly on every compact subset of R^0,

$$(18) \qquad \epsilon^{-1} I_1(\epsilon), \quad \epsilon I_2(\epsilon) \quad \epsilon^{\frac{1}{2}} I_3(\epsilon), \quad I_4(\epsilon) \text{ are } O(1) \text{ as } \epsilon \to 0.$$

It should be mentioned that the assertion of this Lemma remains valid if $\epsilon^{-1} I_1(\epsilon)$, $\epsilon I_2(\epsilon)$ in (18) are replaced by $\epsilon^{-1} I_1(\eta)$, $\epsilon I_2(\delta)$, respectively, where η, δ are functions of ϵ (and x, y) satisfying $0 \leqq \eta \leqq \text{Const.}\, \epsilon$, $0 < \text{const.}\, \epsilon \leqq \delta \leqq y$. This is clear from the monotony of I_1, I_2 with respect to ϵ.

The proof of the Lemma will follow procedures used by Levi; cf. [5], pp. 231-232.

Proof of the Lemma. For $\alpha > 0$, $\beta > 0$ and $b > a \geqq 0$, $c \geqq 0$, put

$$(19) \qquad I_{\alpha\beta}(a, b; c) = \int_a^b t^{-\beta} \int_c^\infty s^\alpha \exp(- s^2/4t) \, ds \, dt.$$

Then the change of variables $s/t^{\frac{1}{2}} = v,\ t = t$ transforms (19) into

$$(20)\qquad I_{\alpha\beta}(a, b; c) = \int_a^b t^{\frac{1}{2}\alpha-\beta+\frac{1}{2}} \int_{c(t)}^{\infty} v^{\alpha} \exp(-v^2/4)\, dv\, dt, \text{ where } c(t) = c/t^{\frac{1}{2}}.$$

Hence, if $\frac{1}{2}\alpha - \beta + 3/2 \neq 0$ and if $a > 0$ when $\frac{1}{2}\alpha - \beta + 3/2 < 0$, then

$$(21)\qquad I_{\alpha\beta}(a, b; c) \leqq I_{\alpha\beta}(a, b; 0) \leqq \text{Const.}\ | b^{\frac{1}{2}\alpha-\beta+3/2} - a^{\frac{1}{2}\alpha-\beta+3/2} |;$$

if $\frac{1}{2}\alpha - \beta + 3/2 = 0$ and $a > 0$, then

$$(22)\qquad I_{\alpha\beta}(a, b; 0) \leqq \text{Const.} \log b/a,$$

where the Const. in (21) and (22) depends only on α and β.

In view of (7) and (7 bis), the integrals (17_1)-(17_4) are majorized by 2 times

$$(23_1)\qquad \int_0^{\epsilon}\int_0^{\infty} \gamma(s, t)\, ds\, dt; \qquad\qquad (23_2)\qquad \int_{\epsilon}^{y}\int_0^{\infty} | \gamma_{tt}(s, t) |\, ds\, dt$$

$$(23_3)\qquad \int_{\epsilon}^{y}\int_0^{\infty} | \gamma_{sss}(s, t) |\, ds\, dt; \qquad\qquad (23_4)\qquad \int_0^{\epsilon}\int_{\delta}^{\infty} | \gamma_t(s, t) |\, ds\, dt,$$

respectively, where $y > \epsilon$. If $t > 0$, then (7) gives

$$(24)\qquad \gamma_t = \tfrac{1}{2}t^{-3/2}(\tfrac{1}{2}s^2/t - 1)\exp(-s^2/4t),$$

$$(25)\qquad 16\gamma_{tt} = (12t^{-5/2} - 12s^2t^{-7/2} + s^4t^{-9/2})\exp(-s^2/4t),$$

$$(26)\qquad 8\gamma_{sss} = 8\gamma_{ts} = (6st^{-5/2} - s^3t^{-7/2})\exp(-s^2/4t).$$

The integral in (23_1) is $I_{0\frac{1}{2}}(0, \epsilon; 0) \leqq \text{Const.}\ \epsilon$. By (25), the function (23_2) is majorized by a constant multiple of

$$I_{0\ 5/2}(\epsilon, y; 0) + I_{2\ 7/2}(\epsilon, y; 0) + I_{4\ 9/2}(\epsilon, y; 0) \leqq \text{Const.}\ | y^{-1} - \epsilon^{-1} |.$$

By (26), (23_3) is majorized by a constant multiple of

$$I_{1\ 5/2}(\epsilon, y; 0) + I_{3\ 7/2}(\epsilon, y; 0) \leqq \text{Const.}\ | y^{-\frac{1}{2}} - \epsilon^{-\frac{1}{2}} |.$$

This proves the statements concerning (17_1)-(17_3).

Since (24) shows that $\gamma_t > 0$ for $0 < t(< \frac{1}{2}\epsilon) \leqq \frac{1}{2}s^2$, the integral in (23_4) is majorized by the sum of

$$\int_{\delta}^{\infty}\int_0^{\frac{1}{2}\epsilon} \gamma_t ds\, dt < \int_0^{\infty} \gamma(s, \tfrac{1}{2}\epsilon)\, ds = 2^{\frac{1}{2}} \int_0^{\infty} \exp(-s^2/4)\, ds$$

and a constant multiple of

$$I_{2\ 5/2}(\tfrac{1}{2}\epsilon, \epsilon; 0) + I_{0\ 3/2}(\tfrac{1}{2}\epsilon, \epsilon; 0) \leqq \text{Const.} \log(\epsilon/\tfrac{1}{2}\epsilon).$$

This proves the assertion concerning (17_4).

7. *Proof of* (ii*). In order to prove that (8) satisfies $\partial^* u = f$ on R^0, it will first be shown that the "Lebesgue constant" $L(h) = L(h; x, y)$ defined by

$$L(h) = \iint_R |\partial_h G(x, y)|\, dx' dy' \tag{27}$$

is bounded as $h \to 0$, if the point (x, y) of R^0 is fixed. Put

$$\epsilon = h^2 > 0, \qquad h > 0. \tag{28}$$

The contribution of the rectangle $|x'| < 1$, $y < y' < 1$ to (27) is 0. The contribution of the rectangle $|x'| < 1$, $y - 2\epsilon < y' < y$ is majorized by the sum of

$$\epsilon^{-1} \int_{y-2\epsilon}^{y} \int_{-1}^{1} |G(x+h, y) + G(x-h, y) - 3G(x, y)|\, dx' dy'$$

and

$$\epsilon^{-1} \int_{y-2\epsilon}^{y-\epsilon} \int_{-1}^{1} G(x, y-\epsilon)\, dx' dy'.$$

In view of the Lemma (cf. (17_1) and (18)), this contribution is $O(1)$ as $\epsilon \to 0$.

The integrand of (27) does not exceed the sum of the absolute values of

$$\epsilon^{-1}\{G(x+h, y) - 2G(x, y) + G(x-h, y)\} - G_{xx}(x, y) \tag{31}$$

and

$$G_y(x, y) - \epsilon^{-1}\{G(x, y) - G(x, y-\epsilon)\}, \tag{32}$$

since $G_{xx} - G_y = 0$. The expressions (31) and (32) can be written as

$$\epsilon^{-1} \int_0^h \int_{-w}^{w} \int_0^v G_{xxx}(x+u, y)\, du\, dv\, dw$$

and

$$\epsilon^{-1} \int_0^{\epsilon} \int_0^v G_{yy}(x, y-u)\, du\, dv, \tag{33}$$

respectively. Consequently, in order to prove the boundedness of (27), it is sufficient to prove the boundedness of

$$(34)\qquad \epsilon^{-1}\int_0^h \int_{-w}^{w} \Big| \int_0^v \Big\{ \int_0^{y-2\epsilon} \int_{-1}^{1} | G_{xxx}(x+u,y) |\, dx'dy' \Big\} du \Big|\, dv\, dw$$

and of

$$(35)\qquad \epsilon^{-1}\int_0^{\epsilon} \int_0^{v} \Big\{ \int_0^{y-2\epsilon} \int_{-1}^{1} | G_{yy}(x,y-u) |\, dx'dy' \Big\} du\, dv.$$

In view of the Lemma (cf. (17_3)), the interior integral $\{\cdot\cdot\cdot\}$ in (34) is $O(\epsilon^{-\frac{3}{2}})$ as $\epsilon \to 0$, uniformly in u. Hence (34) is

$$(36)\qquad \epsilon^{-1}\int_0^h \int_{-w}^{w} \Big| \int_0^v O(\epsilon^{-\frac{3}{2}})\, du \Big|\, dv\, dw = O(\epsilon^{-1}h^3\epsilon^{-\frac{3}{2}}) = O(1),$$

since $h = \epsilon^{\frac{5}{6}}$. Similarly, the interior integral $\{\cdot\cdot\cdot\}$ in (35) is $I_2(2\epsilon; x, y-u)$, by (17_2), and is therefore $O(\epsilon^{-1})$ uniformly for $u (\leqq \epsilon)$, by the Lemma and the remark following it (where $\delta(2\epsilon; x, y-u) = 2\epsilon - u \geqq \frac{1}{2}(2\epsilon)$). Thus (35) is $O(\epsilon^{-1}\epsilon^2\epsilon^{-1}) = O(1)$.

This proves the boundedness of $L(h)$ as $h \to 0$. The assertion (11) can be deduced at once. For let (x_0, y_0) be a point of R^0 and let $\epsilon > 0$. Choose $\delta = \delta_\epsilon > 0$ so small that the square S: $|x - x_0| \leqq \delta$, $|y - y_0| \leqq \delta$ is in R^0 and that $|f(x,y) - f(x_0,y_0)| < \epsilon$ for any point (x,y) of S. Write $f = f_1 + f_2$, where $f_1(x,y)$ is $f(x,y)$ or $f(x_0,y_0)$ according as (x,y) is not or is in S. Thus $|f_2(x,y)| < \epsilon$ at every point (x,y) of R. Let u_1 and u_2 denote the functions (8), when f is replaced by f_1 and f_2, respectively. (The discontinuities of f_1 and f_2 on the boundary of S do not affect the argument to follow.) Since f_1 is smooth (in fact, a constant on S), the function u_1 is of class C^∞ and satisfies $\partial u_1 = f_1 \equiv f(x_0,y_0)$ in the interior of S. In particular, $\partial^* u_1 = f(x_0,y_0)$ in S. Since $|f_2| \leqq \epsilon$, it follows from $L(h) = O(1)$ that, as $h \to 0$, $\limsup |\partial_h u_2| \leqq \text{Const.}\,\epsilon$ for all (x,y) in R^0. Hence, if $(x,y) = (x_0,y_0)$, then $\limsup |\partial_h u - f(x_0,y_0)| \leqq \text{Const.}\,\epsilon$, where $h \to 0$. Since $\epsilon > 0$ is arbitrary, this proves that the limit $\partial^* u(x_0,y_0)$ exists and is $f(x_0,y_0)$, as claimed by (ii*).

8. *Proof of* (iv). The assertions (i* bis) and (ii*) imply that it is sufficient to consider the particular solution (8) of $\partial^* u = f$ in the proof of (iv). It will be shown that, if (x,y) is a point of R^0, the difference between the integral in (15) and the ratio $-2\pi^{\frac{1}{2}}\{u(x,y) - u(x,y-h)\}/h$ tends

to 0 as $\epsilon = 2\,|\,h\,| \to 0$. To this end, the corresponding "Lebesgue constant" will be proved to be bounded.

Let

$$\epsilon = 2\,|\,h\,|, \quad \text{where} \quad h \gtrless 0, \tag{37}$$

and put

$$L_1(h) = \int_0^{y-2|h|} \int_{-1}^{1} |\, G_y(x,y) - h^{-1}\{G(x,y) - G(x,y-h)\}\,|\, dx'dy'. \tag{38}$$

Thus, if $D(\epsilon) = R(y-\epsilon)$, the Lebesgue constant of the problem is majorized by

$$L_1(h) + 2\,|\,h\,|^{-1} I_1(2\,|\,h\,|\,; x, y) + 2\,|\,h\,|^{-1} I_1(2\,|\,h\,| - h\,; x, y-h)\,;$$

cf. (17_1). If the region $D(\epsilon)$ is the portion of $R(y)$ outside the rectangle $|\,x'-x\,| < \epsilon^{\frac{1}{2}}$, $y > y' > y - \epsilon$, then (17_4) must be added to this majorant.

In view of the Lemma, the boundedness of the Lebesgue constant follows if it is shown that (38) is bounded as $h \to 0$. This has been proved in Section 7 when $h > 0$; cf. the considerations concerning (32), (33) and (35). The case $h < 0$ can be treated in the same manner, since the proof of the boundedness of (35) depended on the inequality $u \leqq \epsilon$ (which is satisfied if $u \leqq 0$).

Let (x, y) be fixed and let $\epsilon > 0$ be so small that the points (x', y') of the rectangle S: $|\,x'-x\,| \leqq \frac{1}{2}\epsilon^{\frac{1}{2}}$, $|\,y'-y\,| \leqq \epsilon$ are in R^0. Let f be written as $f = f_1 + f_2$, where $f_1 = f_1(x', y')$ is $f(x', y')$ or $f(x, y)$ according as the point (x', y') of R is not or is in S (so that f_1 is constant on S). Let u_1, u_2 denote the functions (8) which result if f is replaced by f_1, f_2, respectively.

Clearly, u_1 is of class C^∞ on the interior of S; in particular $\partial u_1/\partial y$ and $\partial^2 u_1/\partial x^2$ are given by the formulae (16), where L is the limit in (15 bis) in which f is replaced by f_1; cf. [5], p. 238. Since $f_1(x', y') - f_1(x, y) = 0$ if (x', y') is on S, the corresponding L is L_ϵ, where L_ϵ is defined as in (15 bis). Thus, $\lim \partial u_1/\partial y$ and/or $\lim \partial^2 u_1/\partial x^2$, where $\epsilon \to 0$, exists if and only if the limit in (15 bis) does.

In view of the boundedness of the "Lebesgue constants," the value of

$$\limsup \,|\, \{u_2(x, y+h) - u_2(x, y)\}/h \,| \qquad (h \to 0)$$

does not exceed const. η, where η is the maximum of $|\,f_2\,|$ on R. Hence

$$\limsup \,|\, \{u(x, y+h) - u(x, y)\}/h - \partial u_1/\partial y \,| \leqq \text{const.}\, \eta, \quad \text{where } h \to 0.$$

The continuity of f implies that $\eta = \eta(\epsilon) \to 0$ as $\epsilon \to 0$. Consequently,

$u_y(x,y)$ exists if and only if $\lim \partial u_1/\partial y$, where $\epsilon \to 0$, does (that is, if and only if the limit in (15 bis) does). This proves the first part of (iv) (and the formulae (16) as well, if (15) and/or the limit in (15 bis) exists). The last part of (iv) is a consequence of (16).

Remark. The proof of the boundedness of the "Lebesgue constants" shows that if u is the function (8) or any continuous solution of (11), where f is continuous, then *any one of the following conditions is necessary and sufficient for the existence of* $u_{xx}(x,y)$ *and/or* $u_y(x,y)$ *at the point* (x,y) *of* R^0:

(I) $\lim\,(u(x,y)-u(x,y-h))/h$ exists as $h \to +0$;

(II) $\lim\,(u(x,y)-u(x,y-h))/h$ exists as $h \to -0$;

(III) $\lim\,(u(x+h,y)-2u(x,y)+u(x-h,y))/h^2$ exists as $h \to 0$.

9. *Proof of* (iiib) *and* (iiiβ). An example of a continuous function $f(x,y)$ for which (8) fails to possess the partial derivative u_y at a point of R^0 can now be given at once. Define an auxiliary function $f^* = f^*(s,t)$ as follows: For all s, put

$$f^* = 0 \text{ if } t \leqq 0,\quad f^* = (1-\tfrac{1}{2}s^2/t)/\{(1+\tfrac{1}{2}s^2/t)\log t\} \text{ if } t > 0. \tag{39}$$

Clearly, f^* is continuous. Let (x_0, y_0) be a point of R^0 and put

$$f(x,y) = f^*(x_0 - x, y_0 - y). \tag{40}$$

It will be verified that, for the choice (40) of f, the limit (15) does not exist at $(x,y) = (x_0, y_0)$.

The integral in (15), where $D(\epsilon) = R(y-\epsilon)$, becomes

$$\int_{\epsilon}^{y}\int_{-1}^{1} f^*(s,t)\gamma_t(s,t)\,ds\,dt$$

under the substitution $s = x - x'$, $t = y - y'$. If the integration variables are changed by the transformation $v = s/t^{\frac{1}{2}}$, $t = t$, then (24) and (39) show that the last integral becomes

$$\int\!\!\int \tfrac{1}{2}\,|\,t\log t\,|^{-1}(1-\tfrac{1}{2}v^2)^2(1+\tfrac{1}{2}v^2)^{-1}\exp(-v^2/4)\,dv\,dt,$$

where the domain of integration is $|\,v\,| \leqq t^{-\frac{1}{2}}$, $\epsilon \leqq t \leqq y$. The integrand is non-negative and so the last integral is obviously minorized by

$$\text{const.} \int_{\epsilon}^{y} | \, t \log t \, |^{-1} \, dt, \text{ where const.} > 0.$$

This proves (iiib) and (iiiβ).

10. *The assertions* (ii′) *and* (ii″). The proofs of these assertions will be omitted. For (ii′) can be proved by the procedure employed in [8] dealing with the equation $\Delta u = f$, except that, in the arguments of [8], Petrini's criterion [6], pp. 131-134, is replaced by assertion (iv). A corresponding remark applies to Theorem (ii″), which can be proved by adopting the method of [3], pp. 267-269, which is an elaboration of the one in [8].

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Gevrey, "Sur les équations aux dérivées du type parabolique," *Journal de Mathématiques*, ser. 6, vol. 9 (1913), pp. 305-471.

[2] P. Hartman and A. Wintner, "On the solutions of the equation of heat conduction," *American Journal of Mathematics*, vol. 72 (1950), pp. 367-395.

[3] ——— and A. Wintner, "On the existence of Riemannian manifolds which cannot carry non-constant analytic or harmonic functions in the small," *ibid.*, vol. 75 (1953), pp. 260-276.

[4] E. Holmgren, "Sur l'équation de la propagation de la chaleur," *Arkiv for Matematik, Astronomi och Fysik*, vol. 4 (1908), no. 14.

[5] E. E. Levi, "Sull'equazione del calore," *Annali di Matematica*, ser. 3, vol. 14 (1908), pp. 187-264.

[6] H. Petrini, "Les dérivées premières et secondes du potentiel logarithmique," *Journal de Mathématiques*, ser. 6, vol. 5 (1908), pp. 127-223.

[7] H. Weyl, "Bemerkungen zum Begriff des Differentialquotienten gebrochener Ordnung," *Vierteljahrsschrift der naturforschenden Gesellschaft in Zürich*, vol. 62 (1917), pp. 296-302.

[8] A. Wintner, "On the Hölder restrictions in the theory of partial differential equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 731-738.

[9] S. Zaremba, "Contribution à la théorie d'une équation fonctionelles de la physique," *Rendiconti del Circolo Matematico di Palermo*, vol. 19 (1905), pp. 140-150.

ON ELLIPTIC MONGE-AMPÈRE EQUATIONS.*

By Philip Hartman and Aurel Wintner.

1. *Introduction.* The partial differential equation to be considered will not be the general elliptic Monge-Ampère equation but that of the particular form

(1) $$rt - s^2 = f(x, y, z, p, q),$$

where

(2) $$f > 0$$

(so that the linear terms in r, s, t of the general Monge-Ampère equation are missing).

A classical theorem of S. Bernstein states that a solution $z = z(x, y)$ of class C^3 of an analytic elliptic partial differential equation of second order is analytic. According to Lichtenstein [9], pp. 935-936, the assumption that z is of class C^3 can be lightened to the hypothesis that z is of class C^2, if the differential equation is linear in the second order partial derivatives. It will be shown below that the same is true when the differential equation is of the type (1). The proof will be based on standard techniques in the theory of elliptic partial differential equations and on a manifestation of a general principle, according to which a function $z(x, y)$ cannot be smoother as a function of the variables $(u, v) = (x, q(x, y))$ than is expected a priori.

For other manifestations of the latter principle, see [4], pp. 133-134, and for those of a similar principle applied to the variables $(u, v) = (p(x, y), q(x, y))$, see [5], p. 306 and [6]. Along the lines of the above-mentioned principle, the Legendre-like transformation $(x, y) \to (x, q)$ (cf. [1], p. 140) was applied to $rt - s^2$ in [3], pp. 152-153, in another connection.

The theorem announced above is as follows:

(i) *If $f(x, y, z, p, q)$ is a positive analytic function of its arguments on some five-dimensional domain and if $z = z(x, y)$ is a solution of class C^2 of* (1) *on some (x, y)-domain, then z is analytic.*

This theorem (i) is known in the particular cases where f in (1) depends

* Received May 5, 1953.

only on (x, y) and not on z, p, q (Pogoreloff [11]); so that (1) is of the form

$$rt - s^2 = \phi(x, y) \qquad (\phi > 0) \tag{3}$$

(in which case the assumption on the solution z can be further lightened). (i) and the results to be obtained below will depend on the consideration of equations of the type (3) in which it is assumed that ϕ is of class C^1 or, in fact, subject only to a Hölder condition. It may be mentioned in this connection that Pogoreloff's results [11] cannot be extended from (3) to (1), since his method involves the assumption that ϕ in (3) is of class C^3.

If $0 < \lambda < 1$, a function $\phi(x, y)$ will be called of class $C^0(\lambda)$ on the circle

$$D\colon\ x^2 + y^2 < 1 \tag{4}$$

if, for every positive $r < 1$, there exists a constant $M = M(r)$ satisfying

$$|\phi(x, y) - \phi(x', y')| \leqq M(|x - x'|^\lambda + |y - y'|^\lambda) \tag{5}$$

for every pair of points (x, y), (x', y') of the circle

$$D_r\colon\ x^2 + y^2 < r^2 \qquad (r < 1;\ D_1 = D). \tag{6}$$

If $n = 1, 2, \cdots$ and $0 < \lambda < 1$, a function $\phi(x, y)$ will be called of class $C^n(\lambda)$ on D if it possesses all partial derivatives of the n-th order and these are of class $C^0(\lambda)$ on D. In this terminology, the main result of the present paper can be stated as follows:

(I_n) *Let $\phi(x, y)$ be positive and of class $C^n(\lambda)$ on D for a fixed $n \geqq 0$, and let $z = z(x, y)$ be a solution of class C^2 of (3) on D. Then $z(x, y)$ is of class $C^{n+2}(\mu)$ on D for every $\mu < \lambda$.*

These assertions, where $n = 0, 1, \cdots$, are analogues of standard theorems dealing with solutions of the Poisson equation

$$r + t = \phi(x, y). \tag{7}$$

In addition to their use in the proof of (i), the assertions (I_n) have applications in differential geometry. For example, it will be shown that (I_n) implies the following theorem:

(ii_n) *Let $S\colon z = z(x, y)$ be a surface of class C^2 possessing a positive Gaussian curvature $K(x, y) = (rt - s^2)/(1 + p^2 + q^2)^2$ of class C^n. Then the surface S is of class C^{n+1}.*

The assumption that K is of class C^n can be lightened to the hypothesis that K is of class $C^{n-1}(\lambda)$ and, under this lightened hypothesis, it can be asserted that S is of class $C^{n+1}(\mu)$ for every positive $\mu < \lambda$.

If the assumption of the existence of a positive Gaussian curvature K of class C^n is replaced by the assumption of the existence of a mean curvature H of class C^n, then the corresponding assertion is known; cf. [4], Theorem (i), p. 127. On the other hand, (ii_n) seems to be new for two reasons; first, that it allows $n = 2$ and, second, that S is assumed to be of class C^2 only (rather than of class C^3, even when $n > 2$); cf. [4], Theorem (ii), p. 127. Thus (ii_2) answers in the positive the question formulated in [4], p. 128, since (ii_2) means that not only the first but also *the second of the theorems of* [4], p. 127, *remains true for* $n = 2$.

Pogoreloff [11], p. 88, has a result analogous to (ii_n). It states that if a convex surface S: $z = z(x, y)$ has a positive Gaussian curvature K which, *as a function of the normal vector,* is of class C^n, where $n \geqq 3$, then S is of class C^{n+1}. If the assumption that S is convex is strengthened to the assumption that S is of class C^2, then the corresponding assertion is contained in (ii_{n-1}); cf. the proof of (ii_n) below. Thus the "$n \geqq 3$" in this version of Pogoreloff's theorem can be improved to "$n \geqq 2$" (the case $n = 1$ being correct, but trivial). On the other hand, the theorem of Pogoreloff does not seem to contain (ii_n) even for large n, since his assumption requires that K be smooth as a function of the normal vector (cf. [5], p. 306), that is, of (p, q), rather than, as in (ii_n), of (x, y).

2. *An "integrated" form of* (3). If $z = z(x, y)$ is of class C^2 on D, then the form of Green's theorem given by the Lemma of [2], p. 761, shows that

$$(8) \qquad \int_J p(sdx + tdy) = \iint_B (rt - s^2)\,dxdy = -\int_J q(rdx + sdy),$$

for every domain B bounded by a piecewise smooth Jordan curve J contained in D. Hence (3) can be writen in the form

$$(9) \qquad \int_J (ps - qr)\,dx + (pt - qs)\,dy = 2\iint_B \phi(x, y)\,dxdy.$$

It is clear from (3) that

$$(10) \qquad t = t(x, y) \neq 0$$

on D. Hence the transformation $(x, y) \to (u, v)$ defined by

$$(11) \qquad u = x, \qquad v = q(x, y),$$

a transformation of class C^1 having the Jacobian

$$(12) \qquad \partial(u, v)/\partial(x, y) = t(x, y),$$

is one-to-one if (x, y) is restricted to a small vicinity of a point of D, say to D_ϵ, where $\epsilon > 0$ is sufficiently small. Let D^* denote the (u, v)-image of D_ϵ under the transformation (11).

The equations (11) imply that

$$(13) \qquad du = dx, \quad dv = sdx + tdy \quad \text{or} \quad dx = du, \quad dy = (dv - sdu)/t;$$

hence

$$(14_1) \qquad z_u = p - qs/t; \qquad (14_2) \qquad z_v = q/t.$$

It is easily verified that (13) and (14) give

$$(15) \qquad (ps - qr)dx + (pt - qs)dy = -(rt - s^2)z_v du + z_u dv.$$

Thus, by (3) and (12), the relation (9) can be written as

$$(16) \qquad \int_J \phi z_v du - z_u dv = -2 \int\int_B \phi/t \, dudv,$$

where B is the interior of any piecewise smooth Jordan curve J contained in D^*. The transformation of (3) into (16) was the object of this section.

3. *Functions $\phi(x, y)$ of class C^1.* Under the assumption that ϕ (> 0) is of class C^1, the relation (16), which is an identity in B, implies, by the Lemma of [2], p. 761, that

$$(17) \qquad \int_J \phi^{\frac{1}{2}} z_v du - \phi^{-\frac{1}{2}} z_u dv = \int\int_B \psi \, dudv,$$

where

$$(18) \qquad \psi = \psi(u, v) = -2\phi^{\frac{1}{2}}/t - \tfrac{1}{2}\phi^{-3/2}(z_u\phi_u + \phi z_v\phi_v).$$

Consider the system of partial differential equations

$$(19) \qquad \alpha_u = \phi^{\frac{1}{2}}\beta_v, \qquad \alpha_v = -\phi^{-\frac{1}{2}}\beta_u.$$

Since ϕ is positive and of class C^1, the coefficients $\phi^{\frac{1}{2}}$, $-\phi^{-\frac{1}{2}}$ in (19) are of

class C^1 on D^*. It follows therefore from a theorem of Lichtenstein [10] that (19) has a solution

$$(20) \qquad \alpha = \alpha(u, v), \qquad \beta = \beta(u, v)$$

which is of class $C^1(\mu)$ for every $\mu < 1$. Furthermore, (20) possesses a positive Jacobian

$$(21) \qquad \chi = \partial(\alpha, \beta)/\partial(u, v)$$

and defines a one-to-one mapping of D^* onto some (α, β)-domain D^{**}. It should be remarked, for later reference, that if ϕ, as a function of (u, v), is of class $C^n(\lambda)$, then (20) is of class $C^{n+1}(\mu)$ for every $\mu < \lambda$; cf. [10].

According to (21), the formulation (17) of (3) is transformed by (20) into

$$(22) \qquad \int_J z_\beta d\alpha - z_\alpha d\beta = \int\int_B \psi/\chi \, d\alpha d\beta.$$

Formally, this is equivalent to the Poisson equation

$$(23) \qquad z_{\alpha\alpha} + z_{\beta\beta} = -\psi/\chi.$$

4. *Proof of* (I_n) *when* $n > 0$. Let $z = z(x, y)$ be a solution of class C^2 of (3). It is sufficient to prove that z has the asserted properties of smoothness in a vicinity D_ϵ of $(x, y) = (0, 0)$. It can also be supposed that

$$(24) \qquad q = z_y(x, y) \neq 0$$

at $(0, 0)$, hence on D_ϵ for sufficiently small $\epsilon > 0$; for otherwise $z(x, y)$ could be replaced by $z(x, y) + \text{const.}\, y$.

If ϕ is of class $C^n(\lambda)$, where $n \geqq 1$ and $0 < \lambda < 1$, then ϕ is of class C^1 and, on D_ϵ, the differential equation (3) can be written in the form (22) after the transformations $(x, y) \to (u, v) \to (\alpha, \beta)$, given by (11) and (20).

Since ψ/χ in the double integral (22) is a continuous function of (α, β), it follows from (22) that z as a function of (α, β) is of class $C^1(\mu)$ for every positive $\mu < 1$ (see [8], pp. 96-100); in fact, z as a function of (α, β) differs only by an harmonic function from the logarithmic potential having a density proportional to ψ/χ. Since (20) is of class $C^1(\mu)$ for every $\mu < 1$, it follows that z as a function of (u, v) is of class $C^1(\mu)$ for every $\mu < \lambda$ (< 1). Also, p, q are of class C^1 in terms of (u, v), since they are of class C^1 in (x, y). Hence (14_2) shows that t as a function of (u, v) (hence of (x, y)) is of

class $C^0(\mu)$, where $0 < \mu < \lambda$. Hence (14_1) and (24) show that the same is true of s as a function of (u, v) (hence of (x, y)). Finally, (3) and (10) show that the same is true of r as a function of (u, v) (hence of (x, y)). Consequently, $z = z(x, y)$ is a function of class $C^2(\mu)$ for every $\mu < \lambda$.

Let $n > 0$ be fixed. The assertion (II_n) will be proved by induction. Let $0 \leqq k < n$ and let it be assumed that $z = z(x, y)$ is of class $C^{k+2}(\mu)$ for every $\mu < \lambda$. It will be shown that z is of class $C^{k+3}(\mu)$ for every $\mu < \lambda$.

The transformation (11) is of class $C^{k+1}(\mu)$. Thus z, z_u, z_v, t, as functions of (u, v), are of class $C^{k+1}(\mu)$, $C^k(\mu)$, $C^k(\mu)$, $C^k(\mu)$, respectively. Hence (18) is of class $C^k(\mu)$.

Since $k + 1 \leqq n$ and the transformation (11) is of class $C^{k+1}(\mu)$ for every $\mu < \lambda$, it follows that ϕ as a function of (u, v) on D^* is of class $C^{k+1}(\mu)$ for every $\mu < \lambda$. Consequently, the transformation (20) is of class $C^{k+2}(\mu)$ and so its Jacobian χ is of class $C^{k+1}(\mu)$ for every $\mu < \lambda$; cf. the remark following (21). Since (18) is of class $C^k(\mu)$, the function ψ/χ on the right of (22) is of class $C^k(\mu)$ as a function of (α, β) or of (u, v).

By known facts on Poisson's equation (23) or its integrated form (22) (cf., e. g., [8]), it follows from (22) that z is a function of class $C^{k+2}(\mu)$ for every $\mu < \lambda$ as a function of (α, β). Since the transformation (20) is of class $C^{k+2}(\mu)$, it is seen that z is of class $C^{k+2}(\mu)$ as a function of (u, v).

Hence z_u, z_v (as well as p, q) are of class $C^{k+1}(\mu)$ as functions of (u, v). It follows from (14_1), (14_2), (3), respectively, that t, s, r are of class $C^{k+1}(\mu)$ as functions of (u, v), hence also as functions of (x, y).

Since the second order partial derivatives of z are of class $C^{k+1}(\mu)$, it follows that z is of class $C^{k+3}(\mu)$ for every $\mu < \lambda$. This completes the proof of (I_n) for $n > 0$.

5. *Proof of* (i). Let f and z satisfy the conditions of (i). Put

$$\phi(x, y) = f(x, y, z(x, y), p(x, y), q(x, y)).$$

Then $\phi(x, y)$ is of class C^1 and $z(x, y)$ satisfies (3). It follows from Section **3** and the first part of Section **4**, that $z(x, y)$ is of class $C^2(\mu)$ for every $\mu < 1$.

Hence ϕ is of class $C^1(\mu)$ for every $\mu < 1$. It follows therefore from (I_1) that z is of class $C^3(\mu)$ for every $\mu < 1$. (A simple induction shows that z is of class C^∞.) Hence the assertion (i) follows from Bernstein's theorem.

6. *Proof of* (I_0). Under the assumptions of (I_0), the function $\phi(x,y)$ cannot be assumed to be of class C^1, so that the reduction of (16) to (17) (and then to (22)) is not valid. The assertion of (I_0) will be proved by a device used by E. Hopf [7] to show that the C^3-assumption in Bernstein's theorem can be replaced by a $C^2(\lambda)$-assumption.

In view of (16) and the arguments of Section **4** (used to pass from the $C^k(\mu)$-character of z as a function of (u,v) to its $C^{k+1}(\mu)$-character as a function of (x,y)), the assertion (I_0) will be proved if the following theorem is verified:

If $z(u,v)$, $\phi(u,v)>0$, $\psi(u,v)$ are functions of class C^1, $C^0(\lambda)$, C^0 ($=$ continuous), respectively, on some simply connected (u,v)-domain D^* and if

$$\int_J \phi z_v du - z_u dv = \int\int_B \psi\, dudv \tag{25}$$

holds for every domain B bounded by a piecewise smooth Jordan curve J contained in D^*, then $z(u,v)$ is of class $C^1(\mu)$ on D^* for every $\mu<\lambda$.

It is sufficient to show that $z(u,v)$ is of class $C^1(\mu)$ on every circle contained in D^*. Also, it can be supposed that the closure of the circle D: $u^2+v^2<1$ is contained in D^*, and it is sufficient to show that z is of class $C^1(\mu)$ on D.

Let (u_0,v_0) be a point of D and put

$$\phi_0=\phi(u_0,v_0). \tag{26}$$

Then (25) can be written as

$$\int_J \phi_0 z_v du - z_u dv = \int_J (\phi_0-\psi) z_v du + \int\int_B \psi\, dudv. \tag{27}$$

Also

$$\int_J z_u du + z_v dv = 0 \tag{28}$$

for every J. Introducing complex notation, put

$$w=w(u,v)=\phi_0^{\frac{1}{2}}u+iv \tag{29}$$

and

$$\omega=\omega(u,v)=\phi_0^{\frac{1}{2}}z_v+iz_u. \tag{30}$$

14

Then, if (28) is multiplied by $i\phi_0^{\frac{1}{2}}$ and added to (27), there results the relation

$$\int_J \omega dw = \int_J (\phi_0 - \phi) z_v du + \int\int_B \psi \, dudv. \tag{31}$$

The Lemma in [2], p. 761, implies that if $G = G(u, v)$ is of class C^1 on a domain containing J, then

$$\int_J G\omega dw = \int_J G(\phi_0 - \phi) z_v du$$
$$+ \int\int_B \{G\psi + G_v(\phi_0 - \phi) z_v + \omega(iG_u - \phi_0^{\frac{1}{2}} G_v)\} dudv.$$

Let $w_0 = \phi_0^{\frac{1}{2}} u_0 + iv_0$ and put

$$G = G(u, v) = G(u, v; u_0, v_0) = (w - w_0)^{-1}; \tag{32}$$

so that $iG_u - \phi_0^{\frac{1}{2}} G_v \equiv 0$ for $| w - w_0 | \neq 0$. Thus

$$\int_J G\omega dw = \int_J G(\phi_0 - \phi) z_v du + \int\int_B (G\psi + G_v(\phi_0 - \phi) z_v) dudv, \tag{33}$$

if $B = B(\epsilon)$ is the domain bounded by the circles $u^2 + v^2 = 1$ and $| w - w_0 |^2 \equiv \phi_0 (u - u_0)^2 + (v - v_0)^2 = \epsilon^2$, where $\epsilon > 0$ is small.

As $\epsilon \to 0$, the contribution of the line integral around $| w - w_0 | = \epsilon$ to the left side of (33) tends to $-2\pi i \omega(w_0)$, while the contribution of the corresponding line integral on the right side of (33) tends to 0, since, according to (26),

$$\phi_0 - \phi(u, v) = O(| u - u_0 |^\lambda + | v - v_0 |^\lambda). \tag{34}$$

But (34) also shows that, as $\epsilon \to 0$, the double integral over $B = B(\epsilon)$ in (33) tends to the corresponding (absolutely convergent) integral over D: $u^2 + v^2 \leqq 1$. Hence

$$2\pi i \omega(w_0) = \int_C G\omega dw - \int_C G(\phi_0 - \phi) z_v du \tag{35}$$
$$- \int\int_D (G\psi + G_v(\phi_0 - \phi) z_v) dudv,$$

where D: $u^2 + v^2 < 1$ and C: $u^2 + v^2 = 1$.

In view of (26) and (30),

$$(36) \qquad \omega(w_0) = \phi^{\frac{1}{2}}(u_0, v_0) z_v(u_0, v_0) + i z_u(u_0, v_0).$$

Hence it is sufficient to show that $\omega(w_0) = \omega(u_0, v_0)$ is of class $C^0(\mu)$ on $u_0^2 + v_0^2 < 1$, for every $\mu < \lambda$.

Since $G = \{\phi_0^{\frac{1}{2}}(u_0, v_0)(u - u_0) + i(v - v_0)\}^{-1}$, by (36), it is clear that the functions

$$\int_C G\omega dw, \qquad \int_C G(\phi_0 - \phi) z_v du, \qquad \iint_D G\psi \, du dv$$

of (u_0, v_0) are of class $C^0(\lambda)$ on $u_0^2 + v_0^2 < 1$. Thus in order to prove the case $n = 0$ of (I_n), all that remains to be verified is that

$$\iint_D G_v(\phi_0 - \phi) z_v du dv,$$

as a function of (u_0, v_0), is of class $C^0(\mu)$ on $u_0^2 + v_0^2 < 1$ for every $\mu < \lambda$. But this fact is contained in [7], Hilfssatz 3, pp. 203-204, since it is clear that the kernel $K(X; Z) = K(x_1, x_2; z_1, z_2) = (\phi^{\frac{1}{2}}(x_1, x_2) z_1 + i z_2)^{-1}$ satisfies the conditions of that lemma and that $G = K(X; Y - X)$ if $X = (u_0, v_0)$ and $Y = (u, v)$.

7. *Proof of* (ii_n). Let $(\mathrm{ii}_n{}^*)$ denote the strengthened form of (ii_n), as formulated after the statement of (ii_n). Thus $(\mathrm{ii}_n{}^*)$ claims that if $z = z(x, y)$ is of class C^2 and if the Gaussian curvature $K(x, y)$ is of class $C^n(\lambda)$, where $n = 0, 1, \cdots$ and $0 < \lambda < 1$, then z is of class $C^{n+2}(\mu)$ for every $\mu < \lambda$.

The function $z = z(x, y)$ satisfies (3), where

$$(37) \qquad \phi(x, y) = K(x, y)(1 + p^2(x, y) + q^2(x, y))^2.$$

If K is of class $C^0(\lambda)$, then (37) is also. Hence $(\mathrm{ii}_0{}^*)$ follows from (I_0).

Let $n > 0$ and assume the truth of $(\mathrm{ii}_0{}^*), \cdots, (\mathrm{ii}_{n-1}{}^*)$. If K is of class $C^n(\lambda)$, it follows from $(\mathrm{ii}_{n-1}{}^*)$ that (37) is of class $C^{n+1}(\mu)$ for every $\mu < \lambda$. Hence $(\mathrm{ii}_n{}^*)$ follows from (I_n).

It is worth mentioning that the proof of $(\mathrm{ii}_n{}^*)$, where $n \geqq 1 \neq 0$, depends only on Sections **3** and **4**, and not on the device used in Section **6** in the proof of (I_0).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Darboux, "*Leçons sur la théorie générale des surfaces*," vol. 1, Paris, 1887.

[2] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[3] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[4] ——— and A. Wintner, "On the curvatures of a surface," *ibid.*, vol. 75 (1953), pp. 127-141.

[5] ——— and A. Wintner, "On the third fundamental form of a surface," *ibid.*, vol. 75 (1953), pp. 298-334.

[6] ——— and A. Wintner, "On pieces of convex surfaces," *ibid.*, vol. 75 (1953), pp. 477-487.

[7] E. Hopf, "Über den funktionalen, inbesondere den analytischen Charakter der Lösungen elliptischer Differentialgleichungen zweiter Ordnung," *Mathematische Zeitschrift*, vol. 34 (1932), pp. 194-233.

[8] L. Lichtenstein, "Über einige Integrabilitätsbedingungen zweigliedriger Differentialausdrücke mit einer Anwendung auf den Cauchyschen Integralsatz," *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, vol. 9 (1910), pp. 84-100.

[9] ———, "Über den analytischen Charakter der Lösungen regulärer zweidimensionaler Variationsprobleme," *Bulletin de l'Académie des Sciences de Cracovie*, ser. A (1913), pp. 915-941.

[10] ———, "Zur Theorie der konformen Abbildung. Konforme Abbildung nichtanalytischer singularitätenfreier Flächenstücke auf ebene Gebiete," *ibid.*, ser. A (1916), pp. 192-217.

[11] A. B. Pogoreloff, "Regularity of convex surfaces with given Gaussian curvature," *Matematicheskii Sbornik*, vol. 31 [vol. 73] (1952), pp. 88-103 (in Russian).

SIMPLE DIFFERENTIALS OF SECOND KIND ON HODGE MANIFOLDS.*

By MAXWELL ROSENLICHT.

1. Let V be a complex analytic manifold. A *simple differential* on V is a meromorphic exterior differential form of degree one. The simple differential ω is said to be *of the second kind* if, for each $p \in V$, there exists an open neighborhood U_p of p and a function f_p, meromorphic in U_p, such that $\omega = df_p$ in U_p. A differential of the first kind (one that is closed and everywhere holomorphic) is clearly of the second kind. Similarly, if F is any function meromorphic on all of V, then the differential $\omega = dF$ is of the second kind; such a differential is said to be *exact*.

Let V have complex dimension n and let ω be a simple differential of second kind on V. The polar locus of ω (i. e., the set of points at which ω is not holomorphic) is an analytic subvariety of V of dimension $n-1$ (or empty). If γ is any path on V whose endpoints are not in this polar locus then one can define the integral $\int_\gamma \omega$, which is a complex number depending only on ω, the endpoints of γ, and the homotopy class of γ among paths having these endpoints fixed. In particular, if γ is any 1-cycle, the period $\int_\gamma \omega$ depends only on the homology class of γ. If $\int_\gamma \omega = 0$ for each cycle γ, then we can write $\omega = dF$, where F is meromorphic on all of V. It follows that if Ω_2 denotes the vector space (over the complex number field C) of simple differentials of second kind on V, Ω_e the subspace of exact differentials, and B_1 the first Betti number of V, then

$$\dim \Omega_2/\Omega_e \leq B_1.$$

If V is an algebraic curve Γ (compact Riemann surface) it is easy to see that *equality holds.* For if Γ has genus g, we can find distinct points $P_1, \cdots, P_g \in \Gamma$ such that there exists no nonconstant meromorphic function on Γ with polar divisor at worst $P_1 + \cdots + P_g$; then if $\omega_1, \cdots, \omega_g$ is a basis for the differentials of first kind on Γ and η_i is a normal differential of

* Received April 9, 1953.

second kind with polar divisor $2P_i$ $(i=1,\cdots,g)$, no nontrivial linear combination of $\omega_1,\cdots,\omega_g,\eta_1,\cdots,\eta_g$ is exact, so that $\dim\Omega_2/\Omega_e \geqq 2g = B_1$.

If V is an arbitrary compact complex manifold, equality need not hold, not even if V is Kähler. For example, let V be a multitorus of dimension n and let $\omega_1,\cdots,\omega_n$ be a basis for the invariant holomorphic differential forms of degree one. Then any meromorphic differential form on V of degree one is of the form $F_1\omega_1+\cdots+F_n\omega_n$, where each F_i is a meromorphic function on V. If $n>1$, V may be chosen so that all everywhere meromorphic functions are constant. In this case $\dim\Omega_2/\Omega_e = n$, while $B_1 = 2n$.

However, equality has long been known to hold if V is an algebraic surface, and Kodaira has recently extended this result to algebraic varieties of dimension 3. We shall prove that equality actually holds if V is any compact *Hodge manifold* (i. e. a Kähler manifold in which the basic 2-form is homologous to a scalar multiple of an integral cocycle), which includes all nonsingular algebraic varieties in projective space. The proof we give, a reduction to the known case of algebraic curves by means of the theory of abelian varieties, is quite simple, and possesses the additional virtue of being extendible to algebraic varieties over arbitrary ground fields (in spite of the seemingly crucial role played here by topology), as we intend to show in a later paper.

2. THEOREM. *If V is a compact Hodge manifold, then* $\dim\Omega_2/\Omega_e = B_1$.

Since V is Kähler, B_1 is even, say $B_1=2q$, and the vector space of differentials of first kind is of dimension q. Let $\omega_1,\cdots,\omega_q$ be linearly independent differentials of first kind. Define a map Φ of 1-chains on V into points of C^q, the space of q complex variables, by setting

$$\Phi(\gamma) = \left(\int_\gamma \omega_1,\cdots,\int_\gamma \omega_q\right).$$

If H is the first homology group of V with integral coefficients then the periods $\Phi(H)$ are a discrete subgroup of rank $2q$ of the additive group C^q, so $A = C^q/\Phi(H)$ is a multitorus (compact complex Lie group). Since V is a Hodge manifold, A is actually an *abelian variety*, i. e., A is a multitorus that is analytically homeomorphic to an algebraic variety. We define a map ϕ of V into A by fixing a point $p_0 \in V$ and letting, for any $p \in V$, $\phi(p)$ be the image in A of $\Phi(\gamma)$, where γ is any path in V from p_0 to p. ϕ is clearly well defined and is a complex analytic map of V into A. By its definition, ϕ induces an isomorphism of the rational first homology group of V onto that of A. If η is any meromorphic differential on A, then by applying a suitable group translation on A we can get another meromorphic differential

η' that has the property that the polar locus of η' does not contain $\phi(V)$. If η is of the second kind, so is its translate η', and since η and η' have the same periods they differ by an exact differential. For any η of second kind on A whose polar locus does not contain $\phi(V)$, the differential $\phi^{-1}(\eta)$ is a differential of second kind on V. For any cycle γ on V we have

$$\int_\gamma \phi^{-1}(\eta) = \int_{\phi(\gamma)} \eta,$$

so the periods of $\phi^{-1}(\eta)$ are all zero if and only if the periods of η are all zero, i. e., $\phi^{-1}(\eta)$ is exact if and only if η is exact. Thus if $\eta_1, \cdots, \eta_{2q}$ are differentials of second kind on A that are linearly independent modulo exact differentials, then we can assume that $\phi(V)$ is not contained in the polar locus of any η_i, and then $\phi^{-1}(\eta_1), \cdots, \phi^{-1}(\eta_{2q})$ are differentials of second kind on V that are linearly independent modulo exact differentials. That is, our theorem holds for V if it holds for A. We may thus restrict our attention to abelian varieties.

If A is an abelian variety and $\mathfrak{g}$ is a finite subgroup of A, then $A' = A/\mathfrak{g}$ is also an abelian variety. Let π denote the natural homomorphism of A onto A'. If η is a differential of the second kind on A', then $\pi^{-1}(\eta)$ is of the second kind on A. π induces an isomorphism of the rational first homology group of A onto that of A', so the periods of $\pi^{-1}(\eta)$ are all zero if and only if the periods of η are all zero, and if our theorem holds for A' then it also holds for A. Now if A, A' are *arbitrary* abelian varieties such that there exists a finite subgroup $\mathfrak{g}$ of A such that A' is analytically isomorphic to $A/\mathfrak{g}$, then A and A' are said to be *isogenous*. Isogeny is an equivalence relation. Hence if our theorem holds for A then it also holds for any abelian variety isogenous to A.

Let A_1, A_2 be abelian varieties. Then the direct product $A = A_1 \times A_2$ is also an abelian variety. We shall show that the theorem holds for A if and only if it holds for both A_1 and A_2. Bases for the first homology groups of A_1 and A_2 (homeomorphic to the subspaces $A_1 \times (0)$ and $(0) \times A_2$ respectively of A) together give a homology basis for A. Let π_1, π_2 denote the projections of A onto its first and second factors respectively. If η_1, η_2 are differentials of second kind on A_1, A_2 respectively, then $\eta = \pi_1^{-1}(\eta_1) + \pi_2^{-1}(\eta_2)$ is a differential of second kind on A, and the periods of η are all zero if and only if the periods of η_1 and η_2 both are all zero. It follows that if the theorem holds for both A_1 and A_2, then it also holds for A. Conversely, let η be any differential of second kind on A. Then any translate η' of η on A

is also of the second kind and has the same periods as η. If we choose η' such that the polar locus of η' does not contain $(0)\times(0)$, then this polar locus will contain neither $A_1\times(0)$ nor $(0)\times A_2$. η' then induces differentials η_1, η_2 on $A_1\times(0)$, $(0)\times A_2$ respectively, and η_1, η_2 are of the second kind. The differential $\eta'-\pi_1^{-1}(\eta_1)-\pi_2^{-1}(\eta_2)$ has zero period along each cycle in either $A_1\times(0)$ or $(0)\times A_2$, and hence along each cycle in A. Thus for any differential η of second kind on A there exist differentials η_1, η_2 of second kind on A_1, A_2 respectively such that $\eta-\pi_1^{-1}(\eta_1)-\pi_2^{-1}(\eta_2)$ is exact. As a consequence, if the theorem holds for A it must also hold for both A_1 and A_2.

A *simple abelian variety* is one which has no proper abelian subvarieties. Since every abelian variety is isogenous to a direct product of simple abelian varieties, to prove our theorem in complete generality it suffices to prove it for simple abelian varieties. So let A be simple. Let Γ be an algebraic curve in A and let J be the jacobian variety of Γ. Then one knows that there exists an abelian variety B such that J is isogenous to $A\times B$. By the preceding two paragraphs, to prove our theorem for A it suffices to prove it for J.

Let g be the genus of the algebraic curve Γ and let ϕ be the canonical map of Γ into its jacobian variety J. (If we let Γ be the V of the first paragraph of this proof, then $J=A$, and the present ϕ is the same as the preceding one.) Let Γ^g be the direct product $\Gamma\times\Gamma\times\cdots\times\Gamma$ (g times) and let $\Gamma^{(g)}$ be the g-fold symmetric product of Γ. If ω is a differential of second kind on Γ and, if $P_1\times\cdots\times P_g$ is a general point of Γ^g, then $\omega(P_1)+\cdots+\omega(P_g)$ is clearly a differential of second kind on Γ^g; since it remains unchanged if we permute the points $P_1,\cdots,P_g$, it is actually a closed meromorphic differential form on the symmetric product $\Gamma^{(g)}$. Since $\Gamma^{(g)}$ and J are birationally equivalent, $\omega(P_1)+\cdots+\omega(P_g)$ defines a closed meromorphic differential form on J. If $q\,\varepsilon\,J$, we can write $q=\phi(Q_1)+\cdots+\phi(Q_g)$, where $Q_1,\cdots. Q_g\,\varepsilon\,\Gamma$ and "$+$" refers to the group addition on J; furthermore $Q_1,\cdots,Q_g$ are unique except for order, unless q lies on a certain $(g-2)$-dimensional analytic subvariety W of J. If $q\not\varepsilon W$ and $Q_1,\cdots,Q_g$ are distinct, then the map $P_1\times\cdots\times P_g\to\phi(P_1)+\cdots+\phi(P_g)$ is an analytic homeomorphism of a neighborhood of $Q_1\times\cdots\times Q_g\,\varepsilon\,\Gamma^g$ onto a neighborhood of q; thus $\omega(P_1)+\cdots+\omega(P_g)$ is locally exact in a neighborhood of q on J. Next let $q\,\varepsilon\,J$ be any point not contained in W. Write

$q = \phi(Q_1) + \cdots + \phi(Q_g)$ and let f be a meromorphic function on Γ such that $\omega - df$ is finite at each point $Q_1, \cdots, Q_g$. Then $f(P_1) + \cdots + f(P_g)$ is a meromorphic function on J such that

$$\omega(P_1) + \cdots + \omega(P_g) - d(f(P_1) + \cdots + f(P_g))$$

is a differential on J that is finite in a neighborhood of q, and therefore holomorphic in that neighborhood. As a consequence, $\omega(P_1) + \cdots + \omega(P_g)$ is locally exact at each point of J-W. Finally, let $q \in W$, let U be a small cellular neighborhood of q in J and let $p_0 \in U - U \cap W$ be a point at which $\omega(P_1) + \cdots + \omega(P_g)$ is holomorphic. Then the integral $\int_{p_0}^{p} (\omega(P_1) + \cdots + \omega(P_g))$, taken along a path lying entirely in $U - U \cap W$ is independent of the path (since $U - U \cap W$ is simply connected, W having dimension $g - 2$). Thus we can write $\omega(P_1) + \cdots + \omega(P_g) = dF$ in $U - U \cap W$, where F is meromorphic. By the continuity theorem for meromorphic functions of several complex variables, F is meromorphic in all of U. Therefore $\omega(P_1) + \cdots + \omega(P_g)$ is a differential of the second kind on J. Now note that under the inverse of the map $\phi : \Gamma \to J$, the differential $\omega(P_1) + \cdots + \omega(P_g)$ induces the differential ω on Γ. (If $\phi(\Gamma)$ is contained in the polar locus of $\omega(P_1) + \cdots + \omega(P_g)$ we still get a well defined differential on Γ by applying a constant translation to $\phi(\Gamma)$). Thus if $\omega(P_1) + \cdots + \omega(P_g)$ is exact, then ω is itself exact. It follows that if ω_i $(i = 1, \cdots, 2g)$ are differentials of the second kind on Γ that are linearly independent modulo exact differentials, then the corresponding differentials $\omega_i(P_1) + \cdots + \omega_i(P_g)$ on J are also linearly independent modulo exact differentials. Since the first Betti number of J is $2g$, our theorem holds for J. The proof is complete.

3. It is of interest to note that if ω is a simple differential of second kind on the compact Hodge manifold V, then the singularities of ω are the singularities of exact differentials. More precisely, for each $p \in V$ there exists an everywhere meromorphic function F_p on V such that $\omega - dF_p$ is holomorphic in a neighborhood of p. To prove this, note that it suffices to show that we can find a set of B_1 basic differentials of second kind none of which has a pole at p. The first paragraph of § 2 shows that it suffices to show this if V is an abelian variety. But here we can translate any given set of basic differentials of second kind to get a new such set of differentials none of which has a pole at p.

4. We remark finally that any simple meromorphic differential on a compact Kähler manifold that has exact singularities is of the second kind. More precisely, if V is Kähler and if ω is a meromorphic differential form on V such that for each $p \in V$ there exists a function f_p, meromorphic in a neighborhood of p, such that $\omega - df_p$ is finite in that neighborhood, then ω is automatically closed. For the second degree differential $d\omega$ is of the first kind, so it suffices to prove that all its periods are zero. Let $\sigma = \Sigma_i \alpha_i s_i$ be a singular 2-cycle, where each s_i is a singular simplex and each α_i a constant. To show that $\int_\sigma d\omega = 0$, we may assume that none of the boundaries ∂s_i meets the polar locus of ω and furthermore that each s_i is so small that in a neighborhood U_i of its carrier we have $\omega = \eta_i + df_i$, where η_i is a holomorphic differential in U_i and f_i is a meromorphic function in U_i. Then

$$\int_\sigma d\omega = \Sigma_i \alpha_i \int_{s_i} d\omega = \Sigma_i \alpha_i \int_{s_i} d\eta_i = \Sigma_i \alpha_i \int_{\partial s_i} \eta_i = \Sigma_i \alpha_i \int_{\partial s_i} \omega = \int_{\partial \sigma} \omega = 0,$$

proving our contention. This can be partially generalized to higher degree differentials, but we do not go into this here.

Added in proof. A slightly weaker case of our main result was announced by P. Dolbeault at the second Liège Colloquium on Algebraic Geometry (June 1952).

NORTHWESTERN UNIVERSITY.

NOTE ON FREE MODULAR LATTICES.*

By R. M. THRALL and D. G. DUNCAN.

1. **Introduction.** The free modular lattice with three generators has 28 elements and its structure is well known.[1] The free modular lattice with four generators has infinite dimension and also has lattice-homomorphic images which have finite dimension but infinitely many elements. It is still an open question just where in passing from the three to the four generator case the boundary between finite and infinite occurs. In this note we work from both directions, but there still remains a gap. On the finite side we study two free modular lattices which are extensions of the free modular lattice with three generators. One of these lattices (type A) is of particular interest as a knowledge of its structure is made use of in the theory of representation of algebras.[2] The generating elements for the two types of lattice under consideration are displayed in the following two figures.

On the infinite side we show in section 5 that the free modular lattice generated by a projective root[3] and an additional element has infinite dimension. This includes as a corollary the special case of three generators and the complement of one.

2. **Lattices of Type A.** This type of lattice is generated by a finite chain of "diamond" lattices and a single non-comparable element (m). As an aid in displaying the types of elements which occur in lattices of this type we use the notation t_i to stand for any of x_i, y_i, or z_i, whenever this notation is practical. With this notation we find by constructing the tables of unions and meets that these lattices are composed of the following thirteen (essentially distinct) types of elements.

(1) t_i

(2) m

(3) $t_i \cap m$

(4) $t_i \cup m$

* Received December 2, 1952.

[1] G. Birkhoff, *Lattice Theory* (revised edition, 1948), pp. 68-69.

[2] This application occurs in an as yet unpublished investigation into algebras of bounded representation type by R. Brauer and R. M. Thrall.

[3] A projective root is a lattice $P = \{t; x, y, z; w\}$ such that t is the meet and w the join of each pair selected from (x, y, z).

(5) $t_i \cap (m \cup t_j) \qquad t_i > t_j$

(6) $x_i \cup (m \cap y_i)$

(7) $x_i \cap (m \cup y_i)$

(8) $(x_i \cap m) \cup (y_i \cap m)$

(9) $(x_i \cap m) \cup (y_i \cap m) \cup t_k \qquad x_i > t_k$

(10) $(x_i \cup m) \cap (y_i \cup m)$

(11) $(x_i \cup m) \cap (y_i \cup m) \cap t_k \qquad t_k > x_i$

(12) $[(y_i \cup m) \cap x_i] \cup (m \cap y_i)$

(13) $[(x_i \cup m) \cap y_i] \cup (x_j \cap m) \cup (y_j \cap m) \qquad i < j$

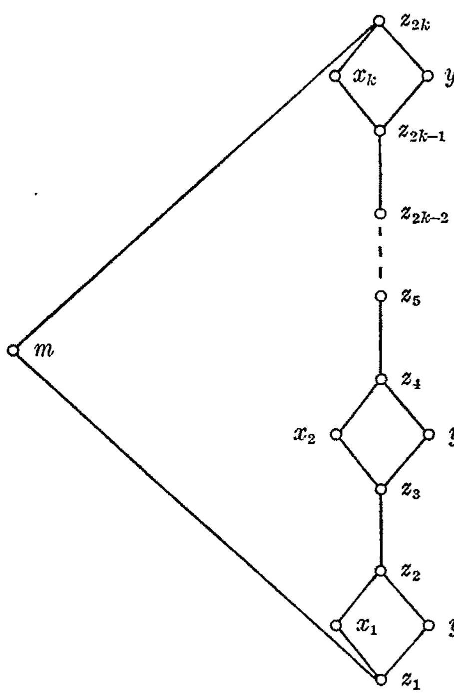

Fig. 1
Type A

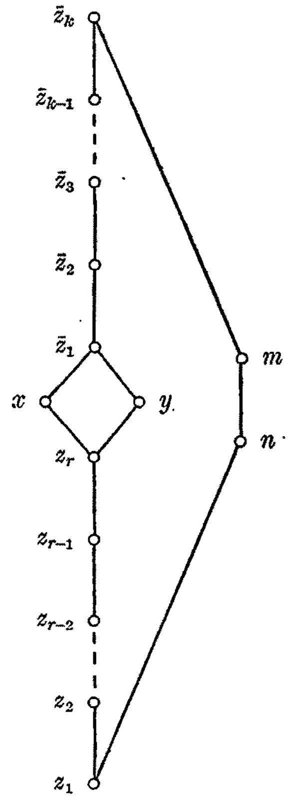

Fig. 2
Type B

Elements (6), (7) and (12) obviously give rise to three essentially analogous elements by interchanges of x_i and y_i. If the generating chain consists of a single diamond we obtain the free modular lattice with three generators.

3. Lattices of Type *B*. For purposes of computation we may think of this type of lattice as being generated by the elements of two lattices of type A, each of which has a single "diamond" embedded in a finite chain and a single non-comparable element; m in one lattice, n in the other. Each generating lattice then consists of twelve types of elements (element 13 of type A is not included since we are dealing here with a single "diamond"). Using the same notation as before we have 23 generating elements (not 24 since the type t_i is common to both generating lattices).

Denoting this set of elements by L, the general lattice of type B is determined by calculating the table of unions $(L \cup L)$ and the table of meets $(L \cap L)$ and then forming all the unions and meets of these elements with each other until no new elements arise.

The lattice $2+1+1$ (Problem 29, G. Birkhoff, *Lattice Theory* (revised edition)) is a lattice of type B with the chains $z_1, \cdots, z_r$ and $\bar{z}_1, \cdots, \bar{z}_k$ both restricted to single elements. This lattice has recently been determined independently by K. Takeuchi who has also constructed the lattice diagram for this lattice of 138 elements (written communication with the authors). Discrepancies occurring between his list of elements and our own are noted in the table of elements given below. This list displays the elements of $2+1+1$ with the corresponding type of element for the general type B. Elements obtainable by duality have been omitted (excepting elements 70 and 71). The eleven types of elements in the general type B which have no counterpart in the restricted case $2+1+1$ are also given.

It should be noted that not every element of the general type B has a dual unless the length r of the chain below the diamond is equal to the length k of the chain above the diamond.

No.	*Element of* $2+1+1$	*Dimension*	*Corresponding element of the general type B*	*Dimension*
1	$z_1 \cap n$	0	$z_i \cap n$	$i-1$
2	$x \cap n$	1	$x \cap n$	r
3	$z_1 \cap m$	1	$z_i \cap m$	i
4	$y \cap n$	1	$y \cap n$	r
5	$(x \cap n) \cup (z_1 \cap m)$	2	$(x \cap n) \cup (z_i \cap m)$	$r+i$
6	z_1	2	z_i	$3i-1$
7	$(x \cap n) \cup (y \cap n)$	2	$(x \cap n) \cup (y \cap n)$	$r+1$
8	$(y \cap n) \cup (z_1 \cap m)$	2	$(y \cap n) \cup (z_i \cap m)$	$r+i$
9	$(y \cup n) \cap m \cap x$	3	$(y \cup n) \cap m \cap x$	$2r+1$
10	$(x \cap n) \cup z_1$	3	$(x \cap n) \cup z_i$	$2i+r$
11	$n \cap [x \cup (m \cap y)]$	3	$n \cap [x \cup (m \cap y)]$	$r+2$
12	$(x \cap n) \cup (y \cap n) \cup (z_1 \cap m)$	3	$(x \cap n) \cup (y \cap n) \cup (z_i \cap m)$	$r+i+1$

No.	Element of $2+1+1$	Dimension	Corresponding element of the general type B	Dimension
13	$(y\cap n)\cup z_1$	3	$(y\cap n)\cup z_i$	$2i+r$
14	$(x\cup n)\cap m\cap y$	3	$(x\cup n)\cap m\cap y$	$2r+1$
15	$x\cap m$	4	$x\cap m$	$2r+2$
16	$x\cap(n\cup y)\cap(m\cup z_1)$	4	$x\cap(n\cup y)\cap(m\cup z_i)$	$2r+i+1$
17	$(y\cup n)\cap[x\cup(n\cap y)]\cap m$	4	$(y\cup n)\cap[x\cup(n\cap y)]\cap m$	$2r+2$
18	$z_2\cap n$	4	$\bar{z}_i\cap n$	$r+i+2$
19	$(z_1\cup n)\cap[(x\cap m)\cup(y\cap m)]$	4	$(z_i\cup n)\cap[(x\cap m)\cup(y\cap m)]$	$r+i+2$
20	$(x\cap n)\cup(y\cap n)\cup z_1$	4	$(x\cap n)\cup(y\cap n)\cup z_i$	$2i+r+1$
21	$(x\cup n)\cap m\cap[y\cup(n\cap x)]$	4	$(x\cup n)\cap[y\cup(n\cap x)]\cap m$	$2r+2$
22	$y\cap(n\cup x)\cap(m\cup z_1)$	4	$y\cap(n\cup x)\cap(m\cup z_i)$	$2r+i+1$
23	$y\cap m$	4	$y\cap m$	$2r+2$
24	$(x\cap m)\cup z_1$	5	$(x\cap m)\cup z_i$	$2r+i+2$
25	$x\cap(n\cup y)$	5	$x\cap(n\cup y)$	$3r+2$
26	$(y\cap n)\cup(x\cap m)$	5	$(y\cap n)\cup(x\cap m)$	$2r+3$
27	$(z_1\cup m)\cap(n\cup y)\cap[x\cup(n\cap y)]$	5	$(z_i\cup m)\cap(n\cup y)\cap[x\cup(n\cap y)]$	$2r+i+2$
28*	$(z_1\cup m)\cap(n\cup x)\cap[y\cup(n\cap x)]$	5	$(z_i\cup m)\cap(n\cup x)\cap[y\cup(n\cap x)]$	$2r+i+2$
29	$(x\cup n)\cap(y\cup n)\cap[(x\cap m)\cup(y\cap m)]$	5	$(x\cup n)\cap(y\cup n)\cap[(x\cap m)\cup(y\cup m)]$	$2r+3$
30	n	5	n	$r+k+3$
31	$(z_2\cap n)\cup(z_1\cap m)$	5	$(\bar{z}_j\cap n)\cup(z_i\cap m)$	$r+i+j+2$
32	$(z_1\cup n)\cap[(m\cap y)\cup x]$	5	$(z_i\cup n)\cap[(m\cap y)\cup x]$	$2i+r+2$
33	$(x\cap n)\cup(y\cap m)$	5	$(x\cap n)\cup(y\cap m)$	$2r+3$
34	$y\cap(n\cup x)$	5	$y\cap(n\cup x)$	$3r+2$
35	$(y\cap m)\cup z_1$	5	$(y\cap m)\cup z_i$	$2r+i+2$
36	$[x\cap(n\cup y)]\cup(x\cap m)$	6	$[x\cap(n\cup y)]\cup(x\cap m)$	$3r+3$
37	$z_1\cup(n\cap y)\cup(m\cap x)$	6	$z_i\cup(n\cap y)\cup(m\cap x)$	$2r+i+3$
38	$(y\cup n)\cap[x\cup(n\cap y)]$	6	$(y\cup n)\cap[x\cup(n\cap y)]$	$3r+3$
39	$(x\cup n)\cap[(x\cap m)\cup(y\cap m)]$	6	$(x\cup n)\cap[(x\cap m)\cup(y\cap m)]$	$2r+4$
40	$(y\cup n)\cap(x\cup n)\cap m\cap z_2$	6	$(y\cup n)\cap(x\cup n)\cap m\cap\bar{z}_i$	$2r+i+3$
41	$(x\cup n)\cap(y\cup n)\cap[(x\cap m)\cup(y\cap m)\cup z_1]$	6	$(x\cup n)\cap(y\cup n)\cap[(x\cap m)\cup(y\cap m)\cup z_i]$	$2r+i+3$
42	$(z_1\cap m)\cup n$	6	$(z_i\cap m)\cup n$	$r+k+i+3$
43	$(z_2\cap n)\cup z_1$	6	$(\bar{z}_j\cap n)\cup z_i$	$r+2i+j+2$
44	$(y\cup n)\cap[(x\cap m)\cup(y\cap m)]$	6	$(y\cup n)\cap[(x\cap m)\cup(y\cap m)]$	$2r+4$
45	$(x\cup n)\cap[y\cup(n\cap x)]$	6	$(x\cup n)\cap[y\cup(n\cap x)]$	$3r+3$
46	$z_1\cup(n\cap x)\cup(m\cap y)$	6	$z_i\cup(n\cap x)\cup(m\cap y)$	$2r+i+3$
47	$[y\cap(n\cup x)]\cup(y\cap m)$	6	$[y\cap(n\cup x)]\cup(y\cap m)$	$3r+3$
48	$x\cap(m\cup y)$	7	$x\cap(m\cup y)$	$3r+4$
49	$(x\cap m)\cup(y\cap n)\cup[x\cap(y\cup n)]$	7	$(x\cap m)\cup(y\cap n)\cup[x\cap(y\cup n)]$	$3r+4$

* This element (No. 28) was missing from Takeuchi's original list and No. 32 of his list was found to be reducible to No. 46 of his list.

No.	Element of $2+1+1$	Dimension	Corresponding element of the general type B	Dimension
50	$(z_2\cap n)\cup(x\cap m)$	7	$(\bar{z}_i\cap n)\cup(x\cap m)$	$2r+i+4$
51	$(x\cup n)\cap(y\cup n)\cap[y\cup(m\cap x)]$	7	$(x\cup n)\cap(y\cup n)\cap[y\cup(m\cap x)]$	$3r+4$
52	$(n\cup x)\cap[(m\cap x)\cup(m\cap y)\cup z_1]$	7	$(n\cup x)\cap[(m\cap x)\cup(m\cap y)\cup z_i]$	$2r+i+4$
53	$(x\cap m)\cup(y\cap m)$	7	$(x\cap m)\cup(y\cap m)$	$2r+5$
54	$(y\cup n)\cap(x\cup n)\cap m$	7	$(y\cup n)\cap(x\cup n)\cap m$	$2r+k+4$
55	$(y\cup n)\cap(x\cup n)\cap(m\cup z_1)\cap z_2$	7	$(y\cup n)\cap(x\cup n)\cap(m\cup z_i)\cap\bar{z}_j$	$2r+i+j+3$
56	$z_1\cup n$	7	$z_i\cup n$	$r+k+2i+3$
57#	$(n\cup y)\cap[(m\cap x)\cup(m\cap y)\cup z_1]$	7	$(n\cup y)\cap[(m\cap x)\cup(m\cap y)\cup z_i]$	$2r+i+4$
58	$(x\cup n)\cap(y\cup n)\cap[x\cup(m\cap y)]$	7	$(x\cup n)\cap(y\cup n)\cap[x\cup(m\cap y)]$	$3r+4$
59	$(z_2\cap n)\cup(y\cap m)$	7	$(\bar{z}_i\cap n)\cup(y\cap m)$	$2r+i+4$
60	$(y\cap m)\cup(x\cap n)\cup[y\cap(x\cup n)]$	7	$(y\cap m)\cup(x\cap n)\cup[y\cap(x\cup n)]$	$3r+4$
61	$y\cap(m\cup x)$	7	$y\cap(m\cup x)$	$3r+4$
62	x	8	x	$3r+5$
63	$[x\cup(n\cap y)]\cap(m\cup y)$	8	$[x\cup(n\cap y)]\cap(m\cup y)$	$3r+5$
64#	$[(m\cap x)\cup n\cup y]\cap[x\cup\{y\cap m\cap(x\cup n)\}]$	8	$[(m\cap x)\cup n\cup y]\cap[x\cup\{y\cap m\cap(x\cup n)\}]$	$3r+5$
65	$z_2\cap(n\cup y)\cap(m\cup z_1)$	8	$\bar{z}_j\cap(n\cup y)\cap(m\cup z_i)$	$2r+i+j+4$
66	$(n\cup x)\cap m$	8	$(n\cup x)\cap m$	$2r+k+5$
67	$(y\cup n)\cap[(m\cap y)\cup x]$	8	$(y\cup n)\cap[(m\cap y)\cup x]$	$3r+5$
68	$(z_2\cap n)\cup(x\cap m)\cup(y\cap m)$	8	$(\bar{z}_i\cap n)\cup(x\cap m)\cup(y\cap m)$	$2r+i+5$
69	$(x\cap m)\cup(y\cap m)\cup z_1$	8	$(x\cap m)\cup(y\cap m)\cup z_i$	$2r+i+5$
70	$(x\cup n)\cap(y\cup n)\cap z_2$ (dual of 69)	8	$(x\cup n)\cap(y\cup n)\cap\bar{z}_i$	$3r+i+4$
71	$(y\cup n)\cap(x\cup n)\cap(m\cup z_1)$ (dual of 68)	8	$(y\cup n)\cap(x\cup n)\cap(m\cup z_i)$	$2r+k+i+4$
72	$(x\cup n)\cap[(m\cap x)\cup y]$	8	$(x\cup n)\cap[(m\cap x)\cup y]$	$3r+5$
73	$(n\cup y)\cap m$	8	$(n\cup y)\cap m$	$2r+k+5$
74	$z_2\cap(n\cup x)\cap(m\cup z_1)$	8	$\bar{z}_j\cap(n\cup x)\cap(m\cup z_i)$	$2r+i+j+4$
75#	$[(m\cap y)\cup n\cup x]\cap[y\cup\{x\cap m\cap(y\cup n)\}]$	8	$[(m\cap y)\cup n\cup x]\cap[y\cup\{x\cap m\cap(y\cup n)\}]$	$3r+5$
76	$[y\cup(n\cap x)]\cap(m\cup x)$	8	$[y\cup(n\cap x)]\cap(m\cup x)$	$3r+5$
77	y	8	y	$3r+5$

The following eleven types of elements occur in the general lattice of type B but not in the restricted case $2+1+1$.

No.	Element of the general type B.		Dimension
78	$(z_i\cap n)\cup(z_j\cap m)$	$i>j$	$i+j-1$
79	$z_i\cup(z_j\cap n)$	$j>i$	$2i+j-1$
80	$z_i\cup(z_j\cap m)$	$j>i$	$2i+j$
81	$(z_i\cap n)\cup(z_j\cap m)\cup z_s$	$i>j>s$	$i+j+s-1$
82	$(x\cap n)\cup(z_i\cap m)\cup z_j$	$i>j$	$r+i+j$
83	$(y\cap n)\cup(z_i\cap m)\cup z_j$	$i>j$	$r+i+j$

These elements (Nos. 57, 64, 75) were missing from our original list.

No.	*Element of the general type B.*		*Dimension*
84	$(y\cap n)\cup(x\cap n)\cup(z_i\cap m)\cup z_j$	$i>j$	$r+i+j+1$
85	$[x\cup(y\cap m)]\cap(z_i\cup n)\cap(z_j\cup m)$	$i>j$	$r+i+j+2$
86	$[y\cup(x\cap m)]\cap(z_i\cup n)\cap(z_j\cup m)$	$i>j$	$r+i+j+2$
87	$(z_i\cup n)\cap(z_j\cup m)\cap\bar{z}_s$	$i>j$	$r+i+j+s+2$
88	$(z_i\cup n)\cap(z_j\cup m)$	$i>j$	$r+k+i+j+3$

4. The free modular lattice generated by $3+1+1$. This lattice may be considered as being generated by three lattices of the type $2+1+1$ each of which consists of 138 distinct elements. The number of generating elements is then calculated to be 334; this is somewhat less than $3(138)=414$ as there is some duplication. For this number of generating elements the method of direct calculation would not appear to be very fruitful. Apparently the structure of the general case $n+1+1$ will have to be obtained by other methods. Indeed, at present, it is not known for which value n (if any) this lattice becomes infinite. A lattice is said to be of type C if it is generated by a finite diamond lattice and a finite chain. Clearly type C includes types A and B as well as $n+1+1$. It is still an open question as to whether there exists an infinite lattice of type C.

5. An infinite lattice. *Let* $P_1=(t;x,y,z;w)$ and $P_2=(t;x,y,u;w)$ be projective roots. Let L be the free modular lattice generated by P_1 and P_2. We shall show that L has finite dimension by exhibiting lattice-homomorphic images of arbitrarily high dimension.

Let r be any positive integer, let W be the free abelian group with $2r$ generators $\alpha_1,\cdots,\alpha_r,\ \beta_1,\cdots,\beta_r$, let T be the subgroup of W consisting of 0 alone, let X, Y, Z, U be the subgroups generated by $\{\beta_1,\cdots,\beta_r\}$, $\{\alpha_1+\beta_1,\cdots,\alpha_r+\beta_r\}$, $\{\alpha_1,\cdots,\alpha_r\}$, $\{\alpha_1,\alpha_2+\beta_1,\cdots\ \alpha_r+\beta_{r-1}\}$ respectively. Then clearly $\{T;X,Y,Z;W\}$ and $\{T;X,Y,U;W\}$ are projective roots. Let L_{2r} be the lattice generated by X, Y, Z, U. We shall show that L_{2r} has dimension $2r$.

Let $R_1=Z\cap U$ and let $R_\nu=(((R_{\nu-1}\cup Y)\cap X)\cup U)\cap Z$ $(\nu=2,\cdots,r)$. It is easy to see that R_ν is the free group generated by $\{\alpha_1,\cdots,\alpha_\nu\}$, $(\nu=1,\cdots,r)$. Next let $R_{r+\nu}=((R_\nu\cup Y)\cap X)\cup Z$ $(\nu=1,\cdots,r)$. Then $R_{r+\nu}$ is generated by $\{\alpha_1,\cdots,\alpha_r,\beta_1,\cdots,\beta_\nu\}$ $(\nu=1,\cdots,r)$. The chain $T\subset R_1\subset\cdots\subset R_{2r}=W$ shows that L_{2r} has dimension $2r$, as claimed.

Now L_{2r} is clearly the lattice-homomorphic image of L under the mapping $x\to X$, $y\to Y$, $z\to Z$, $u\to U$, and so L has infinite dimension.

UNIVERSITY OF MICHIGAN AND
UNIVERSITY OF ARIZONA.

NILPOTENT CHARACTERISTIC SUBGROUPS OF FINITE GROUPS.*

By Reinhold Baer.

Every student of the theory of finite groups is familiar with two nilpotent characteristic subgroups: center and hypercenter. Not so well known, though no less important, are Frattini's and Fitting's subgroups. The former, usually denoted by $\phi(G)$, is the intersection of all the maximal subgroups of G; and the latter, denoted by $F(G)$, is the product of all the normal nilpotent subgroups [and is itself nilpotent]. They are quite closely connected, since it may be shown that $F[G/\phi(G)] = F(G)/\phi(G)$ is an elementary abelian group; see Gaschütz [1] or below § 3, Theorem 1.

Hypercenter and Frattini subgroup are distinguished among nilpotent subgroups by a very powerful property: weak hypercentrality. This property may be stated in various equivalent ways [§ 1, Proposition 2] of which we mention the following one: The normal subgroup N of G is weakly hypercentral, if it is nilpotent and if, for every normal subgroup M of G which contains N, the totality of elements x in M, satisfying $x^{[M:N]^i} = 1$ for some i, is a subgroup. Since products of weakly hypercentral normal subgroups need not be weakly hypercentral, it is necessary to survey their totality. It has the following closure property: If M and N are normal subgroups of G such that $M \leq N$, then weak hypercentrality of N in G is necessary and sufficient for weak hypercentrality of M in G and of N/M in G/M [§ 1, Proposition 3]. If we define, as seems sensible, the weak hypercenter of G as the intersection $H_w(G)$ of all the maximal weakly hypercentral normal subgroups of G, then hypercenter, weak hypercenter and Frattini subgroup of $G/H_w(G)$ equal 1 so that $H_w(G)$ contains both the hypercenter and the Frattini subgroup [§ 2, Theorem 1 and § 1, Corollaries 4, 5].

We make use of these results to prove the following theorem which appears to admit of various applications: If M is a minimal normal subgroup of the group G, if $Z(M < G)$ is the centralizer of M in G and if $G/Z(M < G)$ contains a normal subgroup, not 1, whose order is prime to the order of M, then M is abelian and there exists a subgroup S of G such that $G = SZ(M \leq G)$ and $1 = M \cap S$. The hypothesis concerning the automorphism group $G/Z(M < G)$ of M is indispensable; it is satisfied, for

* Received Sept. 5, 1952.

instance, whenever $G/Z(M < G)$ contains a soluble normal subgroup different from 1 [§ 4, Proposition 2 and § 5, Lemma 2].

It is natural that these results will prove useful in the study of soluble groups. Consequently we obtain in § 5 a great number of properties characteristic of solubility; and in §§ 6, 7 our results are used to continue the study of n-soluble and n-nilpotent groups which we initiated elsewhere.

Notations. We shall consider finite groups only and the order of the group G [of the group element g] shall be denoted by $o(G)$ [by $o(g)$]. It will be convenient to call an element x an *n-element,* if $x^{n^i} = 1$ for some integer i; in other words: if every prime divisor of $o(x)$ is a factor of n. Similarly we shall say that x is a *Pn-element,* if $o(x)$ is prime to n. A group will be termed an *n-group* [a *Pn-group*], if all its elements are n-elements [Pn-elements].

$$Z(G) = \text{center of } G.$$

$$Z(A < G) = \text{centralizer of } A \text{ in } G.$$

$$H(G) = \text{hypercenter of } G.$$

The hypercenter of G may be defined in various equivalent ways. It is, for instance, the intersection of all the normal subgroups N of G which satisfy $Z(G/N) = 1$; and it is also the terminal member of the ascending center chain. Subsets of the hypercenter shall be referred to as hypercentral subsets. For the elementary properties of these concepts see, for instance, Baer [5; section 5].

$$[G, G] = \text{commutator subgroup of } G.$$

The subgroup S of G is a *complement* of the normal subgroup N of G, if $G = NS$ and $1 = N \cap S$.

1. The weakly hypercentral subgroups. We begin by recalling some important concepts. If the totality of n-elements [Pn-elements] in the group G forms a subgroup of G, then we term this characteristic subgroup of G the *n-component* G_n [the *Pn-component* G'_n] of G. If the group G is in particular nilpotent, then these components exist for every n and G is their direct product, see, for instance, Zassenhaus [1; p. 107, Satz 11]. The subgroup S of G is termed a *Pn-complement of G,* if S is an n-group and $[G:S]$ is prime to n. One verifies easily that a subgroup T of G is the n-component of G if, and only if, T is a normal subgroup and a Pn-complement of G.

Like the components these complements need not exist; but whereas components are unique, there may exist several Pn-complements.

LEMMA 1. *Assume that the nilpotent normal subgroup M of G is part of the normal subgroup N of G and that j is a multiple of $[N:M]$. Then*

(a) *there exists a Pj-complement of N and*

(b) $G = TM'_j$ *for every normalizer T of a Pj-complement of N.*

Proof. M'_j is a normal subgroup of G, since it is a characteristic subgroup of the normal subgroup M of G. Since N/M and $M/M'_j \simeq M_j$ are j-groups whereas M'_j is a Pj-group, $o(M'_j)$ and $[N:M'_j]$ are relatively prime. Consequently we may apply Schur's Theorem asserting the existence of a subgroup S of N such that $N = SM'_j$ and $1 = S \cap M'_j$; see, for instance, Zassenhaus [1; p. 125, Satz 25]. Then $S \simeq N/M'_j$ and one sees easily that S is a Pj-complement of N.

If U is any Pj-complement of N, then $o(U) = [N:M'_j]$ and this implies that U too is a complement of M'_j in N since $o(N) = o(U)o(M'_j)$ and $(o(U), o(M'_j)) = 1$. Denote by T the normalizer of U in G. If g is an element in G, then g transforms the normal subgroups N and M'_j into themselves. Hence

$$N = g^{-1}UM'_jg = [g^{-1}Ug]M'_j, \quad 1 = g^{-1}(U \cap M'_j)g = [g^{-1}Ug] \cap M'_j.$$

Thus $g^{-1}Ug$ is also a complement of M'_j in N. Since M'_j is soluble as a subgroup of the nilpotent group M, we may apply the Theorem of Witt-Zassenhaus; see Zassenhaus [1; p. 126, Satz 27]. Hence U and $g^{-1}Ug$ are conjugate in N; and there exists therefore an element w in N such that $w^{-1}Uw = g^{-1}Ug$. From $N = UM'_j$ we deduce the existence of elements u and t in U and M'_j respectively such that $w = ut$. Consequently

$$g^{-1}Ug = w^{-1}Uw = t^{-1}u^{-1}Uut = t^{-1}Ut$$

so that gt^{-1} belongs to the normalizer T of U. The element g belongs therefore to $Tt \leq TM'_j$. Hence $G = TM'_j$; and this completes the proof.

DEFINITION. *The normal subgroup N of G is weakly hypercentral, if it has the following property:*

(W) *If N is part of the normal subgroup M of G, if x and y are elements in N and M respectively, and if $o(x)$ is prime to $[M:N]$ and to $o(y)$, then $xy = yx$.*

A justification of the term "weakly hypercentral" may be found in the following simple and important result.

PROPOSITION 1. *Hypercentral normal subgroups are weakly hypercentral and weakly hypercentral normal subgroups are nilpotent.*

Proof. The normal subgroup N of G is hypercentral if, and only if, every element in N commutes with every element of relatively prime order in G; see, for instance, Baer [5, § 5, Theorem 3]. It is clear that this condition implies (W) and that therefore hypercentrality implies weak hypercentrality. If the normal subgroup N of G is weakly hypercentral and if we let $M = N$ in (W), then it follows that elements of relatively prime order in N commute; and it is well known that this condition is necessary and sufficient for nilpotency of N; see, for instance, Baer [1] or Zassenhaus [1].

PROPOSITION 2. *The following properties of the normal subgroup N of G are equivalent.*

(i) *N is weakly hypercentral.*

(ii) *N is nilpotent; and if N is part of the normal subgroup M of G, then the totality of $[M:N]$-elements in M is a subgroup of M.*

(iii) *If N is part of the normal subgroup M of G, and if j is a multiple of $[M:N]$, then M is the direct product of its j-component M_j and its Pj-component M'_j—note that $M'_j \leq N$.*

Proof. Assume first that N is weakly hypercentral. Then N is nilpotent by Proposition 1. Suppose now that N is part of the normal subgroup M of G and that $n = [M:N]$. We deduce from Lemma 1 the existence of a Pn-complement S of M; and if T is the normalizer of S in G, then $G = TN'_n$. Since every element in N'_n is a Pn-element and every element in S is an n-element, we deduce from the weak hypercentrality of N that every element in N'_n commutes with every element in S. Hence $N'_n \leq T$ and we deduce $G = T$ from $G = TN'_n$. Consequently S is a normal subgroup of G and hence of M; and this implies that the Pn-complement S of M is the n-component of M. Thus we have shown that (ii) is a consequence of (i).

Assume next the validity of (ii). Suppose that the normal subgroup M of G contains N and that j is a multiple of $[M:N]$. Since N is nilpotent, N is the direct product of its j-component N_j and its Pj-component N'_j. These are characteristic subgroups of the normal subgroup N of G; and as such they are normal subgroups of G. Next it follows from (ii) that the

totality V of $[M:N]$-elements in M is a characteristic subgroup of M. Since the normal subgroups N_j and V of M are both j-groups—here we use the fact that j is a multiple of $[M:N]$—their product VN_j is a normal j-subgroup of M. Since V contains every $[M:N]$-element in M, we have $M = NV = N'_j(N_jV)$. Hence M is the product of the normal Pj-group N'_j and the normal j-group N_jV; and now it is clear that M is the direct product of a Pj-group and a j-group. But these are necessarily components of M; and thus we see that (iii) is a consequence of (ii).

That finally (i) is a consequence of (iii), is almost obvious. This completes the proof.

COROLLARY 1. *Suppose that the weakly hypercentral normal subgroup M of G is part of the normal subgroup N of G. Then the k-elements in N form a subgroup of N if, and only if, the k-elements of N/M form a subgroup of N/M.*

Proof. The totality of k-elements in N is a subgroup of N if, and only if, the product of any two k-elements in N is a k-element in N. But this property is clearly invariant under homomorphisms, since k-elements in N/M may be represented by k-elements in N. The totality of k-elements in N/M is therefore certainly a subgroup of N/M whenever the totality of k-elements in N is a subgroup of N.

Assume conversely that the totality of k-elements in N/M is a subgroup K/M of N/M. It is clear that K/M is a characteristic subgroup of the normal subgroup N/M of G/M. Hence K/M is a normal subgroup of G/M and K is a normal subgroup of G. Since K/M is a k-group, $[K:M]$ is a divisor of a suitable power j of k. Since M is weakly hypercentral, we may apply Proposition 2, (iii). Hence K is the direct product of a j-group J and a Pj-group J'. Since J is the j-component of K, J is also the k-component of K. Since every k-element in N represents a k-element in N/M, and since the latter elements belong to K/M, it follows that K contains every k-element in N. Hence J is the k-component of N; and we have shown that the totality of k-elements in N is a subgroup of N, namely J. This completes the proof.

COROLLARY 2. *Supose that the weakly hypercentral normal subgroup M of G is part of the normal subgroup N of G. Then N is nilpotent if, and only if, N/M is nilpotent.*

Proof. A group H is nilpotent if, and only if, the p-elements in H

form, for every prime p, a subgroup of H; see, for instance, Zassenhaus [1]. But it follows from Corollary 1 and the weak hypercentrality of M that the p-elements in N form a subgroup if, and only if, the p-elements in N/M form a subgroup. Hence nilpotency of N and of N/M are equivalent properties.

We note that Corollary 2 generalizes a result due to Gaschütz [1; Satz 10].

COROLLARY 3. *If N is a normal subgroup of G, and if $NZ(N < G)$ is a weakly hypercentral normal subgroup of G, then $G/NZ(N < G)$ does not contain normal subgroups different from 1 whose order is prime to $o(N)$.*

Proof. Assume that the normal subgroup M of G contains $NZ(N < G)$ and that $n = [M : NZ(N < G)]$ is prime to $o(N)$. It follows from the weak hypercentrality of $NZ(N < G)$ and Proposition 2 that M is the direct product of an n-group M_n and a Pn-group M'_n. Since M'_n is the Pn-component of M, and since N is a Pn-subgroup of M, we have $N \leq M'_n \leq NZ(N < G)$. From the first inequality we infer that M_n is part of $Z(N < G)$. Hence $M \leq NZ(N < G)$ implying $M = NZ(N < G)$ and $n = 1$, as we wanted to show.

Remark. If in particular N is a normal p-subgroup of G such that $NZ(N < G)$ is weakly hypercentral, then it follows from Corollary 3 that the order of every normal subgroup, not 1, of $G/NZ(N < G)$ is divisible by p. If $M/NZ(N < G)$ is a nilpotent normal subgroup of $G/NZ(N < G)$, then it is the direct product of a p-group and a Pp-group both of which are normal. It follows that the latter is 1, proving that nilpotent normal subgroups of $G/NZ(N < G)$ are always p-groups.

PROPOSITION 3. *If M and N are normal subgroups of G and $M \leq N$, then weak hypercentrality of N [in G] is necessary and sufficient for weak hypercentrality of M [in G] and of N/M [in G/M].*

Proof. Assume first that N is a weakly hypercentral normal subgroup of G. Consider a normal subgroup K of G which contains M and elements x and y in M and K respectively such that $o(x)$ is prime to $[K:M]$ and $o(y)$. From $KN/N \simeq K/(N \cap K) \simeq [K/M]/[(N \cap K)/M]$ we deduce that $[KN:N]$ is a factor of $[K:M]$. Hence $o(x)$ is also prime to $[KN:N]$. But the normal subgroup KN of G contains the weakly hypercentral normal subgroup N of G; x belongs to M and hence to N and y belongs to K and hence to KN. Application of condition (W) shows $xy = yx$; and this proves weak hypercentrality of M in G. Consider next a normal subgroup H/M of G/M which

contains N/M and consider elements U and V in N/M and H/M respectively such that $o(U)$ is prime to $[H/M:N/M]$ and $o(V)$. There exist elements u and v such that $U = Mu$, $V = Mv$ and such that u is an $o(U)$-element, v is an $o(V)$-element. Since $[H/M:N/M] = [H:N]$, $o(u)$ is prime to $[H:N]$ and $o(v)$; and we may deduce $uv = vu$ from the weak hypercentrality of N in G. But then $UV = VU$; and this proves the weak hypercentrality of N/M in G/M.

Assume conversely that M is a weakly hypercentral subgroup of G and that N/M is a weakly hypercentral subgroup of G/M. Consider a normal subgroup K of G which contains N and let $k = [K:N]$. Since the normal subgroup K/M of G/M contains the weakly hypercentral normal subgroup N/M of G/M, and since $[K/M:N/M] = [K:N] = k$, we deduce from Proposition 2, (ii) that the totality of k-elements in K/M is a characteristic subgroup H/M of K/M and hence a normal subgroup of G/M. Thus H is a normal subgroup of G which contains the weakly hypercentral normal subgroup M of G; and it follows from Proposition 2, (ii) that the totality of $[H:M]$-elements in H is a subgroup T of H. Since H/M is a k-group, $[H:M]$-elements are k-elements. Hence T is a k-group. If furthermore z is a k-element in K, then Mz is a k-element in K/M and belongs therefore to H/M. Hence z is a k-element in H. To show that z is also an $[H:M]$-element, and therefore in T, we prove that k is a factor of $[H:M]$, as follows. From our choices of H and T it follows [by Proposition 2, (iii)] that $K/M = (H/M)(N/M)$ or $K = HN$ and $H = TM$. Hence $K = TN$ and consequently $k = [K:N] = [T:(T \cap N)]$ is a factor of $[T:(T \cap M)] = [H:M]$, as we wanted to show. This completes the proof of the fact that T is the totality of k-elements in K; and thus we have verified the validity of Proposition 2, (ii). Hence N is weakly hypercentral, as we wanted to show. This completes the proof.

PROPOSITION 4. *The normal subgroup N of G is weakly hypercentral if, and only if, N is a nilpotent group whose primary components are weakly hypercentral.*

The necessity of our conditions is an immediate consequence of Propositions 1 and 3 and their sufficiency will be an immediate consequence of the following fact.

LEMMA 2. *If M and N are weakly hypercentral normal subgroups of G, and if $o(M)$ and $o(N)$ are relatively prime, then MN is weakly hypercentral.*

Proof. Consider a normal subgroup Q of G which contains MN and suppose that x and y are elements in MN and Q respectively such that $o(x)$ is prime to $[Q:MN]$ and to $o(y)$. Since x is in MN, and since $o(M)$ and $o(N)$ are relatively prime, there exist uniquely determined elements m and n in M and N respectively such that $x = mn$. We note that MN is the direct product of M and N and that $o(x) = o(m)o(n)$. From

$$[Q:M] = [Q:MN][MN:M] = [Q:MN]o(N)$$

we infer that $o(m)$ is prime to $[Q:M]$, since $o(M)$ is prime to $o(N)$ and since $o(m)$ as a divisor of $o(x)$ is prime to $[Q:MN]$. Since $o(x)$ is prime to $o(y)$, $o(m)$ is prime to $o(y)$ too; and now we infer from the weak hypercentrality of M that $my = ym$. Similarly we see that $ny = yn$. Hence $xy = yx$; and this proves that MN is weakly hypercentral. This completes the proof of Lemma 2 and, as we mentioned before, of Proposition 4.

Example 1. Denote by p an odd prime and by D the direct product of two cyclic groups of order p. The group G arises from D by adjunction of an element s satisfying

$$s^2 = 1,\ s^{-1}ds = d^{-1} \text{ for every } d \text{ in } D.$$

One verifies easily that every cyclic subgroup of D is a weakly hypercentral subgroup of G whereas D itself is not weakly hypercentral. This example shows that Lemma 2 ceases to be true if we omit from it the hypothesis that $o(M)$ and $o(N)$ are relatively prime.

Example 1 shows furthermore that there may exist several maximal weakly hypercentral normal subgroups and that the product of all weakly hypercentral normal subgroups need not be weakly hypercentral. Thus we are led to the following definition: *The weak hypercenter $H_w(G)$ of G* is the intersection of all the maximal weakly hypercentral normal subgroups of G.

PROPOSITION 5. *The following two properties of the normal subgroup N of G are equivalena.*

(i) $N \leq H_w(G)$.

(ii) *If M is a weakly hypercentral normal subgroup of G, then MN is weakly hypercentral.*

Proof. Assume first that $N \leq H_w(G)$ and that M is a weakly hypercentral normal subgroup of G. There exists a maximal weakly hypercentral

normal subgroup R of G which contains M. Then $N \leq H_w(G) \leq R$ by the definition of $H_w(G)$. The normal subgroup MN is therefore part of the weakly hypercentral normal subgroup R of G; and MN is weakly hypercentral by Proposition 3. Thus (ii) is a consequence of (i).

Assume conversely that N meets requirement (ii) and that U is a maximal weakly hypercentral normal subgroup of G. We deduce from (ii) that NU is a weakly hypercentral normal subgroup of G; and this implies $U = NU$ or $N \leq U$ because of the maximality of U. Hence N is part of the intersection $H_w(G)$ of all the maximal weakly hypercentral normal subgroups of G. Thus (i) is a consequence of (ii).

Remark. It might be worth pointing out that the preceding argument is purely lattice theoretical and is applicable in the following situation: a lattice $\wedge$ and subset Σ of $\wedge$ which has the following two properties.

(a) If σ is in Σ, then every part of σ belongs to Σ.

(b) Every element in Σ is contained in some maximal element in Σ.

COROLLARY 4. $H(G) \leq H_w(G)$.

Proof. If N is a weakly hypercentral normal subgroup of G, then $NH(G)/N \leq H(G/N)$ and it follows from Propositions 1 and 3 that $NH(G)/N$ is a weakly hypercentral normal subgroup of G/N. Now we deduce from the weak hypercentrality of N and from Proposition 3 that $NH(G)$ is weakly hypercentral. Thus condition (ii) of Proposition 5 is satisfied by $H(G)$, proving $H(G) \leq H_w(G)$.

PROPOSITION 6. *If the normal subgroup N of G is part of $H_w(G)$, then $H_w(G/N) = H_w(G)/N$.*

Proof. One deduces easily from Proposition 3 that M/N is a maximal weakly hypercentral normal subgroup of G/N if, and only if, M is a maximal weakly hypercentral normal subgroup of G; and from this fact $H_w(G/N) = H_w(G)/N$ is an easy consequence since N is part of every maximal weakly hypercentral normal subgroup of G.

COROLLARY 5. $H_w[G/H_w(G)] = H[G/H_w(G)] = 1$.

This is a simple consequence of Corollary 4 and Proposition 6.

2. Frattini's subgroup. Frattini's subgroup is *the intersection $\phi(G)$ of all the maximal subgroups of the group G.* Since finite groups, not 1,

possess maximal subgroups, $1 < G$ implies $\phi(G) < G$; and it is well known that this characteristic subgroup $\phi(G)$ is always nilpotent; see, for instance, Zassenhaus [1; p. 115].

PROPOSITION 1. *The following two properties of the subset S of the group G are equivalent.*

(i) $S \leq \phi(G)$.

(ii) *If T is a subgroup of G and $G = \{S, T\}$, then $T = G$.*

The simple proof of this well known property may be left to the reader; see Zassenhaus [1; p. 45].

For a further analysis we need some criteria concerned with the complements of minimal normal subgroups. We recall that the subgroup S of G is termed a *complement* of the normal subgroup N of G, if $G = NS$ and $1 = N \cap S$.

LEMMA 1. *The following properties of the abelian minimal normal subgroup M of G and the subgroup S of G are equivalent.*

(i) *S is a complement of M in G.*

(ii) *S is a maximal subgroup of G and $M \not\leq S$.*

(iii) *$G = MS$ and $S \neq G$.*

(iv) *$G = MS$ and $M \not\leq S$.*

Proof. We assume first that $G = MS$ and $1 = M \cap S$. Then $M \neq 1$ is certainly not part of S. Consider now a subgroup T of G such that $S < T$. Then there exists an element t in T which does not belong to S. From $G = MS$ we deduce the existence of elements r and s in M and S respectively such that $t = rs$. From $S < T$ it follows that r is in $M \cap T$; and $r \neq 1$, since otherwise t would be in S. Since M is a minimal normal subgroup of G, M is generated by the elements conjugate to r in G. Since M is abelian and $G = MS$, every element conjugate to r in G is obtained by transforming r by elements in S. Consequently $M \leq \{S, r\} \leq T$. Hence $G = MS = T$ proving that S is a maximal subgroup of G. Thus (ii) is a consequence of (i).

It is clear that (ii) implies (iii) and that (iii) implies (iv).

Assume finally the validity of (iv) and let $J = M \cap S$. Then every element in M transforms J into itself, since M is abelian; and every element in S transforms J into itself, since M is a normal subgroup of G and since

therefore J is a normal subgroup of S. From $G = MS$ we deduce now that J is a normal subgroup of G. Since M is not part of S, $J < M$; and now we deduce $J = 1$ from the minimality of M. Hence S is a complement of M in G; and this completes the proof.

Remark 1. Without the hypothesis, that M be abelian, Lemma 1 ceases to be true as may be seen from easily constructed examples.

Remark 2. If the minimal normal subgroup M of G is contained in every maximal subgroup of G [does not satisfy the preceding condition (ii) for any S], then $M \leq \phi(G)$. Since $\phi(G)$ is nilpotent and $M \neq 1$, it follows that $M \cap Z[\phi(G)] \neq 1$. But $Z[\phi(G)]$ is a characteristic subgroup of a characteristic subgroup of G. Hence $Z[\phi(G)]$ is a characteristic subgroup and $M \cap Z[\phi(G)]$ is a normal subgroup of G. Now it follows from the minimality of M that $M \cap Z[\phi(G)] = M$ or $M \leq Z[\phi(G)]$; and this implies in particular the commutativity of M.

Remark 3. If the minimal normal subgroup M of G is abelian, then M is an abelian group without proper characteristic subgroups. If the prime number p divides the order of M, then M^p is a characteristic subgroup of M and $M^p < M$. Hence $M^p = 1$. It follows that the product of all abelian minimal normal subgroups of G is an abelian group the orders of whose elements are squarefree; and we recall that such an abelian group is called *elementary*.

PROPOSITION 2. *The normal subgroup N of G is the Frattini subgroup of G if, and only if, N has the following properties:*

(a) *If S is a subgroup of G and $G = NS$, then $S = G$.*

(b) *Every abelian minimal normal subgroup of G/N possesses a complement in G/N.*

Remark 4. This result is closely related to Gaschütz [1; Satz 14].

Proof. That $\phi(G)$ has property (a), is a consequence of Proposition 1. Let $G^* = G/\phi(G)$. Then it is clear that $\phi(G^*) = 1$. If M is an abelian minimal normal subgroup of G^*, then we infer from $\phi(G^*) = 1$ and $M \neq 1$ the existence of a maximal subgroup T of G^* which does not contain M. It follows from Lemma 1 that T is a complement of M in G^*. Thus $\phi(G)$ has properties (a) and (b).

Assume conversely that the normal subgroup N of G has properties (a)

and (b). Then we deduce $N \leq \phi(G)$ from (a) and Proposition 1. We let next $G^* = G/N$ and assume by way of contradiction that $\phi(G^*) \neq 1$. Since $\phi(G^*)$ is nilpotent, its center $Z[\phi(G^*)]$ is not 1 either. But the center of the Frattini subgroup is a characteristic subgroup. Consequently there exists a minimal normal subgroup M of G^* which is part of $Z[\phi(G^*)]$. Clearly M is abelian and we deduce from (b) the existence of a complement S of M in G^*. From $MS = G^*$ and $M \leq \phi(G^*)$ we deduce $S = G^*$ [by Proposition 1]. But then $1 = M \cap S = M \cap G^* = M$, an impossibility. Hence $\phi(G^*) = 1$. Consequently N is the intersection of all those maximal subgroups of G which contain N and this implies clearly $\phi(G) \leq N$. This completes the proof of $N = \phi(G)$.

THEOREM 1. $\phi(G) \leq H_w(G)$.

Proof. We recall that $\phi(G)$ is nilpotent. Hence $\phi(G)$ is the direct product of its primary components ϕ_p. We prove first the following fact.

(1) If P is a weakly hypercentral normal p-subgroup of G, then $P\phi_p$ is weakly hypercentral.

Since P and ϕ_p are normal p-subgroups of G, $P\phi_p$ is likewise a normal p-subgroup of G. We want to verify the validity of condition (ii) of § 1, Proposition 2. Consider therefore a normal subgroup N of G which contains $P\phi_p$. If $[N:P\phi_p]$ is not prime to p, then every element in N is an $[N:P\phi_p]$-element and there is nothing to prove. We assume therefore that $k = [N:P\phi_p]$ is prime to p. Since p-groups are nilpotent, we may apply § 1, Lemma 1. Consequently there exists a Pk-complement K of N; and if H is the normalizer of K in G, then $G = HP\phi_p$. Since ϕ_p is part of $\phi(G)$, we may apply Proposition 1. Hence $G = HP$.

Now we let $KP = U$. If x is an element in P, then $U = x^{-1}Ux$, since P is part of U. If y is an element in H, then $P = y^{-1}Py$, since P is a normal subgroup of G; and $K = y^{-1}Ky$, since H is the normalizer of K. Hence $U = y^{-1}Uy$ for y in H. From $G = HP$ it follows now that every element in G transforms U into itself; in other words: $U = KP$ is a normal subgroup of G.

Since K is a Pk-complement of N, K is a k-group whose index $[N:K]$ is prime to k. We recall next that $k = [N:P\phi_p]$ is prime to p and that $P\phi_p$ is a p-group. Consequently K is a complement of $P\phi_p$ in N and in particular K is a Pp-group.

From $U/P \simeq K/(K \cap P) = K$ we deduce that $[U:P]$ is prime to p.

Next we recall that P is weakly hypercentral. It follows therefore from § 1, Proposition 2, (ii) that the totality of $[U:P]$-elements in U is a subgroup V of U. But P is a p-group and $[U:P]$ is prime to p. Hence V is the totality of Pp-elements in U. Since the Pp-group K is a complement of P in U, one verifies that $K = V$ is the totality of Pp-elements in U. Hence K is a characteristic subgroup of the normal subgroup U of G; and this proves that K is a normal subgroup of G.

We recall now that the normal subgroup K of G is a complement of $P\phi_p$ in N. Thus N is the direct product of the p-group $P\phi_p$ and the Pp-group K. It is clear now that K is the totality of $[N:P\phi_p]$-elements in N. Thus we have verified the validity of condition (ii) of § 1, Proposition 2; and this completes the proof of (1).

(2) If N is a weakly hypercentral normal subgroup of G, then $N\phi(G)$ is weakly hypercentral.

From the weak hypercentrality of N we deduce the nilpotency of N [§ 1, Proposition 1]. Thus N is the direct product of its primary components N_p; and it follows from § 1, Proposition 2 that every N_p is weakly hypercentral. We deduce from (1) that $N_p\phi_p$ is weakly hypercentral. It is clear that $N\phi(G)$ is a nilpotent normal subgroup of G with p-component $N_p\phi_p$. Application of § 1, Proposition 4 proves now the weak hypercentrality of $N\phi(G)$.

(2) is equivalent to § 1, Proposition 5, (ii). Thus $\phi(G) \leq H_w(G)$ is a consequence of § 1, Proposition 5.

Corollary 1. *The Frattini subgroup is weakly hypercentral.*

This is an immediate consequence of Theorem 1 and § 1, Proposition 3.

Corollary 2. *G is an n-group if, and only if, $G/\phi(G)$ is an n-group.*

Proof. It is clear that $G/\phi(G)$ is an n-group whenever G is an n-group. Assume conversely that $G/\phi(G)$ is an n-group. By Corollary 1, $\phi(G)$ is weakly hypercentral; and hence it follows from § 1, Proposition 2 that G is the direct product of a $[G:\phi(G)]$-group N and a $P[G:\phi(G)]$-group N'. It is clear that $N' \leq \phi(G)$ and $G = NN'$; and now it follows from Proposition 1 that $G = N$ is a $[G:\phi(G)]$-group. Since $G/\phi(G)$ is an n-group, G itself is an n-group, as we wanted to show.

Corollary 3. *Suppose that m and n are relatively prime integers.*

Then G is the direct product of an m-group and an n-group if, and only if, $G/\phi(G)$ is the direct product of an m-group and an n-group.

Proof. We note first that the group H is the direct product of an m-group and an n-group if, and only if, H satisfies the following three conditions:

(a) H is an mn-group;

(b) the m-elements in H form a subgroup;

(c) the n-elements in H form a subgroup.

By Corollary 2, (a) is satisfied by G if, and only if, (a) is satisfied by $G/\phi(G)$. By Corollary 1, $\phi(G)$ is weakly hypercentral; and consequently we deduce from § 1, Corollary 1 that (b) and (c) are satisfied by G if, and only if, (b) and (c) are satisfied by $G/\phi(G)$. Of these various equivalences Corollary 3 is an immediate consequence.

3. Fitting's subgroup. Following Fitting and Wendt we denote by *$F(G)$ the product of all the nilpotent normal subgroups of G.*

PROPOSITION 1. *$F(G)$ is nilpotent.*

Proof. Since the Sylow subgroups of nilpotent subgroups are themselves nilpotent normal subgroups, and since nilpotent groups are direct products of their Sylow subgroups, it is clear that $F(G)$ is the product of all the normal subgroups of prime power order. But the product of normal p-subgroups is itself a normal p-subgroup. $F(G)$ is therefore the direct product of p-groups [for various primes p] and as such $F(G)$ is nilpotent.

For another proof of this fact see Fitting [1; p. 102, Satz 14].

PROPOSITION 2. *$Z[F(G) < G]/Z[F(G)]$ does not contain soluble normal subgroups different from 1.*

Proof. We note first that

$$Z[F(G)] = F(G) \cap Z[F(G) < G] \leq Z(Z[F(G) < G]).$$

Assume by way of contradiction that $W = Z[F(G) < G]/Z[F(G)]$ contains soluble normal subgroups different from 1. Then we form the product of all the soluble normal subgroups of W which is itself a soluble characteristic subgroup C of W. From $C \neq 1$ and the solubility of C we deduce the existence of an abelian characteristic subgroup $A \neq 1$ of C. Then

$A = B/Z[F(G)]$ where $Z[F(G)] < B \leq Z[F(G) < G]$ and where B is a characteristic subgroup of G. Furthermore $Z[F(G)] \leq Z(B)$ and $B/Z[F(G)]$ is abelian. Hence B is nilpotent so that $B \leq F(G) \cap Z[F(G) < G] = Z[F(G)]$, a contradiction.—For another proof of this fact see Fitting [1; p. 105, Hilfssatz 12].

THEOREM 1. *$F[G/\phi(G)] = F(G)/\phi(G)$ is an elementary abelian group, so that $\phi[F(G)] \leq \phi(G)$.*

Proof. Let $F[G/\phi(G)] = W/\phi(G)$. It is clear that $\phi(G)$ is part of $F(G)$ and that $F(G)/\phi(G)$ is nilpotent. Hence $F(G) \leq W$. We deduce from § 2, Corollary 1 that $\phi(G)$ is weakly hypercentral; and we deduce now from § 1, Corollary 2 that the nilpotency of W is a consequence of the nilpotency of $W/\phi(G) = F[G/\phi(G)]$. Thus $W \leq F(G)$ completing the proof of $W = F(G)$ or of $F[G/\phi(G)] = F(G)/\phi(G)$.

Now we let $G^* = G/\phi(G)$. Then $\phi(G^*) = 1$. Assume by way of contradiction that $F(G^*)$ is not abelian. Then the commutator subgroup $C = [F(G^*), F(G^*)]$ of $F(G^*)$ is different from 1. The intersection D of C and the center $Z[F(G^*)]$ of the nilpotent group $F(G^*)$ is therefore likewise different from 1. Since C and $Z[F(G^*)]$ are characteristic subgroups of $F(G^*)$, D is likewise a characteristic subgroup of the characteristic subgroup $F(G^*)$. Hence D is a characteristic subgroup, not 1, of G^*. Consequently D contains a minimal normal subgroup M of G^*. It is clear that M is abelian, since M is part of the center of $F(G^*)$. From $\phi(G^*) = 1$ we infer the existence of a maximal subgroup S of G^* which does not contain M; and it follows from § 2, Lemma 1 that S is a complement of M in G^*. From $M \leq F(G^*)$ and Dedekind's Law we infer now that

$$F(G^*) = G^* \cap F(G^*) = MS \cap F(G^*) = M[S \cap F(G^*)].$$

Since M is contained in the center of $F(G^*)$, every element in M commutes with every element in $S \cap F(G^*)$; and it follows now that $F(G^*)$ is the direct product of the abelian group M and of $S \cap F(G^*)$. The commutator subgroup C of $F(G^*)$ is consequently contained in $S \cap F(G^*)$. We recall now that $M \leq D \leq C$ and that S is a complement of M in G^*. Hence $M = M \cap C \leq M \cap S = 1$, a contradiction which proves the commutativity of $F(G^*)$.

As an abelian group $F(G^*)$ is the direct product of its primary components. If P is the p-component of $F(G^*)$, then P^p is a characteristic subgroup of the characteristic subgroup P of the characteristic subgroup $F(G^*)$ of G^*. Hence P^p is a characteristic subgroup of G^*. Assume now by

way of contradiction that $P^p \neq 1$. Then P^p contains a minimal normal subgroup Q of G^* which is necessarily abelian. From $\phi(G^*) = 1$ we deduce the existence of a maximal subgroup R of G^* which does not contain Q. It follows from § 2, Lemma 1 that R is a complement of Q in G^*. Since Q is part of the abelian group P, we deduce from Dedekind's Law that P is the direct product of Q and $P \cap R$. Hence $Q = Q \cap P^p = Q^p$, an impossibility since Q is a finite abelian p-group. Thus $P^p = 1$; and this shows that $F(G^*) = F[G/\phi(G)]$ is an elementary abelian group.

Since the ϕ-group of an elementary abelian group is 1, we find that

$$1 = \phi(F[G/\phi(G)]) = \phi[F(G)/\phi(G)] \text{ or } \phi[F(G)] \leq \phi\,(G);$$

and this completes the proof.

Corollary 1. *The following properties of the normal subgroup N of G are equivalent.*

(i) *N is nilpotent.*

(ii) *$N/[N \cap \phi(G)]$ is an elementary abelian group.*

(iii) *$N/[N \cap \phi(G)]$ is an abelian group.*

(iv) *$N/[N \cap \phi(G)]$ is a nilpotent group.*

Proof. If N is nilpotent, then N is part of $F(G)$ [by definition]. Hence

$$N/[N \cap \phi(G)] \simeq N\phi(G)/\phi(G) \leq F(G)/\phi(G);$$

and now it follows from Theorem 1 that (ii) is a consequence of (i).

It is obvious that (ii) implies (iii) and that (iii) implies (iv).

Assume finally the validity of (iv). It follows from § 2, Corollary 1 and § 1, Proposition 3 that $N \cap \phi(G)$ is a weakly hypercentral normal subgroup of G; and consequently we may deduce from (iv) and § 1, Corollary 2 that N is nilpotent. Thus (i) is a consequence of (iv); and this completes the proof.

Remark 1. If we let in Corollary 1 in particular $N = G$, then we see that Wielandt's characterization of nilpotent groups is a special case of the preceding results; see, for instance, Zassenhaus [1; p. 108].

Remark 2. If the normal subgroup N of G satisfies $N \cap \phi(G) = 1$, then it follows from Corollary 1 that N is nilpotent if, and only if, N is an elementary abelian group.

Remark 3. Theorem 1 and Corollary 1 have also been obtained by Gaschütz [1] where different proofs of these and related results may be found.

COROLLARY 2. *If $F(G)$ is weakly hypercentral, then $G/F(G)$ does not contain soluble normal subgroups different from* 1.

Proof. Assume by way of contradiction that $F(G)$ is weakly hypercentral and that $G/F(G)$ contains a soluble normal subgroup different from 1. Then there exists a normal subgroup N of G such that $F(G) < N$ and such that $N/F(G)$ is abelian. Now we may deduce from § 1, Corollary 2 that N is nilpotent; and this is impossible, since it would imply $N \leq F(G) < N$.

Remark 4. The following fairly obvious consequence of Corollary 2 may be worth mentioning.

The group G is nilpotent if, and only if, G is soluble and $F(G)$ is weakly hypercentral.

Combining § 1, Proposition 1 and § 1, Corollary 4, § 2, Theorem 1 we obtain the following inequalities:

$$\phi(G)H(G) \leq H_w(G) \leq F(G).$$

If $F(G)$ happens to be weakly hypercentral, then $F(G)$ is the one and only one maximal weakly hypercentral normal subgroup; and now we deduce from Corollary 2 that, in general, $F(G)$ and $H_w(G)$ will be different. Furthermore it is not difficult to construct groups whose center and Frattini group equal 1, though $H_w(G)$ is different from 1. Thus $H_w(G)$ may be "somewhere" between the limits given above. But it follows from Theorem 1 that $F(G)/\phi(G)$ is an elementary abelian group, showing that the above limits are not "too far apart."

4. The automorphisms of the minimal normal subgroups. We begin by proving the following simple and important fact.

PROPOSITION 1. *$F(G)$ is part of the centralizer of every minimal normal subgroup of G.*

Proof. If M is a minimal normal subgroup of G, and if $M \cap F(G) = 1$, then every element in M commutes with every element in $F(G)$, since their commutators belong to their intersection which is 1. If next $M \cap F(G) \neq 1$, then it follows from the minimality of M that $M = M \cap F(G)$ or $M \leq F(G)$. But $F(G)$ is nilpotent and M is a normal subgroup of G. Hence $M \cap Z[F(G)] \neq 1$. As a characteristic subgroup of a characteristic subgroup $Z[F(G)]$ is a characteristic subgroup of G. Now it follows from the minimality of M that $M = M \cap Z[F(G)]$ or $M \leq Z[F(G)]$. Thus we have seen in either case that $F(G)$ is part of the centralizer of M.

If N is a normal subgroup of G, then the centralizer $Z(N < G)$ of N in G is likewise a normal subgroup of G and $G/Z(N < G)$ is essentially the same as the group of automorphisms of N which are induced in N by inner automorphisms of G. If S is a subgroup of G such that $G = SZ(N < G)$, then every automorphism of N which is induced by an inner automorphism of G is also induced in N by an element of S; and for this reason we shall say that *G is represented in its normal subgroup N by its subgroup S* whenever $G = SZ(N < G)$. There always exist subgroups of G which represent G in N, for instance G itself. Consequently there exist also minimal subgroups representing G in N; and these will be of particular interest to us.

LEMMA 1. *Assume that M is a minimal normal subgroup of G and that G is represented in M by its subgroup S. Then*

(a) *$M \leq S$ or $M \cap S = 1$;*

(b) *$S \cap Z(M < G)$ is a normal subgroup of MS;*

(c) *M is a minimal normal subgroup of MS.*

Proof. If $M \cap S \neq 1$, then $M \cap S$ contains an element $x \neq 1$. But the same elements are conjugate to x in G and in S, since $G = SZ(M < G)$. Consequently $M \cap S$ contains a full set of elements conjugate to x in G. Since M is a minimal normal subgroup of G, M is generated by any one of its elements, not 1, together with its conjugates; and now it is clear that $M \cap S \neq 1$ implies $M \leq S$. By essentially the same argument one sees that M is a minimal normal subgroup of MS.

$S \cap Z(M < G)$ is a normal subgroup of S, since $Z(M < G)$ is a normal subgroup of G. Thus elements in S transform $S \cap Z(M < G)$ into itself. Elements in M commute with every element in $S \cap Z(M < G)$; and now it is clear that $S \cap Z(M < G)$ is a normal subgroup of MS.

LEMMA 2. *Assume that the minimal normal subgroup M of G is abelian and that S is a minimal subgroup representing G in M. Then*

(a) *MS is a minimal subgroup T with the properties: $M \leq T$ and $G = TZ(M < G)$;*

(b) *$S \cap Z(M < G) \leq \phi(S)$;*

(c) *$F(MS) = M[Z(M < G) \cap S] = Z(M < G) \cap MS$.*

Proof. It is clear that $M \leq MS$; and from the commutativity of M we infer $G = SZ(M < G) = [SM]Z(M < G)$. Assume next that Q is a subgroup with the following properties:

$$M \leq Q \leq MS \text{ and } G = QZ(M < G).$$

Then it follows from Dedekind's Law and the normality of M that $Q = M(Q \cap S)$. From $M \leq Z(M < G)$ we deduce now that

$$G = QZ(M < G) = (Q \cap S)MZ(M < G) = (Q \cap S)Z(M < G).$$

Hence $Q \cap S$ represents G in M; and now it follows from the minimality of S that $S = Q \cap S$ or $S \leq Q$. Hence $MS \leq Q \leq MS$ or $Q = MS$ proving the desired minimality of MS.

$J = S \cap Z(M < G)$ is a normal subgroup of S. Suppose now that H is some subgroup of S which satisfies $JH = S$. Then we have

$$HZ(M < G) = HJZ(M < G) = SZ(M < G) = G,$$

since J is part of $Z(M < G)$ and a normal subgroup of S. Hence H represents G in M and we deduce $H = S$ from the minimality of S. Thus J is a normal subgroup of S which satisfies condition (ii) of § 2, Proposition 1. It follows that $J \leq \phi(S)$.

The [by Lemma 1, (b)] normal subgroup J of MS is nilpotent as a subgroup of $\phi(S)$. The normal subgroup M of MS is nilpotent as an abelian group. Hence $MJ \leq F(MS)$. It follows from Lemma 1, (c) that M is a minimal normal subgroup of MS; and now it follows from Proposition 1 that

$$F(MS) \leq Z(M < MS) \leq Z(M < G).$$

Since M is abelian, $M \leq Z(M < G)$. Now we deduce from the three inequalities just derived and Dedekind's Law that

$$MJ \leq F(MS) \leq MS \cap Z(M < G) = M[S \cap Z(M < G)] = MJ;$$

and this shows the validity of the equation (c).

PROPOSITION 2. *If M is a minimal normal subgroup of G such that $G/Z(M < G)$ contains a normal subgroup, not 1, whose order is prime to the order of M, then*

(a) *M is abelian;*

(b) *every minimal subgroup S representing G in M satisfies $S \cap M = 1$ and $S \cap Z(M < G) = \phi(SM)$;*

(c) *the two minimal subgroups H and K representing G in M satisfy $MH = MK$ if, and only if, there exists an element x in M such that $H = x^{-1}Kx$.*

Proof. There exists by hypothesis a normal subgroup Q of G such that $Z(M < G) < Q$ and $[Q : Z(M < G)]$ is prime to the order of M.

Assume first by way of contradiction that $M \cap Z(M < G) = 1$. Then the order of M equals $[MZ(M < G) : Z(M < G)]$ and this number is prime to $[Q : Z(M < G)]$. The groups $MZ(M < G)/Z(M < G)$ and $Q/Z(M < G)$ have consequently relatively prime order so that their intersection is 1. Hence it follows from Dedekind's Law that

$$Z(M < G) = Q \cap MZ(M < G) = Z(M < G)[Q \cap M] \text{ or } Q \cap M \leq Z(M < G);$$

and consequently we find that $Q \cap M = Q \cap M \cap Z(M < G) = 1$, since $M \cap Z(M < G)$ is supposed to equal 1. But $Q \cap M = 1$ implies $Q \leq Z(M < G)$ contradicting $Z(M < G) < Q$. Our assumption $Z(M < G) \cap M = 1$ has therefore led us to a contradiction. Hence $Z(M < G) \cap M \neq 1$; and it follows from the minimality of M that $Z(M < G) \cap M = M$ or $M \leq Z(M < G)$; and this implies the commutativity of M.

Since M is an abelian minimal normal subgroup, there exists a uniquely determined prime number p such that $M^p = 1$ [§ 2, Remark 3].

Consider now some minimal subgroup S representing G in M. Then $G = SZ(M < G)$ and $J = S \cap Z(M < G)$ is a normal subgroup of MS [by Lemma 1, (b)] such that $J \leq \phi(S)$ [by Lemma 2, (b)]. We note furthermore that $N = Q \cap S$ is a normal subgroup of S which contains J. It follows from Dedekind's Law that

$$NZ(M < G) = [Q \cap S]Z(M < G) = Q \cap SZ(M < G) = Q \cap G = Q;$$

and now we deduce from the Isomorphism Law that

$$\begin{aligned} N/J = (S \cap Q)/(S \cap Z(M < G)) &\simeq (S \cap Q)Z(M < G)/Z(M < G) \\ &= Q/Z(M < G). \end{aligned}$$

Thus $[N : J] \neq 1$ is prime to p, since the order of M is a power of p. Denote now by r the greatest divisor of the order of N which is prime to p. Since $[N : J]$ is prime to p, r is a multiple of $[N : J]$. From $J < \phi(S)$ we deduce the weak hypercentrality of the normal subgroup J of S [by § 2, Corollary 1 and § 1, Proposition 3]. Since N is a normal subgroup of S, we deduce from § 1, Proposition 2 that N is the direct product of an r-group N_r and a Pr-group N'_r. Since p is the only prime divisor of the order of N which is prime to r, N'_r is a p-group; and since $[N : J]$ is prime to p, we have $N'_r \leq J$. Clearly N'_r is the totality of p-elements in N and N_r is the totality of Pp-elements in N.

Assume now by way of contradiction that $M \cap S \neq 1$. Then it follows from Lemma 1, (a) that $M \leq S$. Since M is abelian, we have $M \leq Z(M < G)$;

and thus we have shown that $M \leq J$. Since N'_r is the totality of p-elements in N, we find that

$$M \leq N'_r \leq J \leq Z(M < G).$$

That N_r is part of the centralizer of M, follows from the fact that elements in N_r and in N'_r commute and that M is part of N'_r. Hence the direct product N of N_r and N'_r is part of the centralizer of M so that

$$N = N \cap Z(M < G) = S \cap Q \cap Z(M < G) = S \cap Z(M < G) = J.$$

But this contradicts the fact that $[N:J] \neq 1$ which we verified before. Thus we have been led to a contradiction which proves that $M \cap S = 1$.

Next we analyze the normal subgroup $MS \cap Q = M[S \cap Q] = MN$ of MS. We recall that N is the direct product of the p-group N'_r and the Pp-group N_r; and that $N'_r \leq J \leq Z(M < G)$. It follows that every element in N'_r commutes with every element in M and with every element in N_r. Hence M and N'_r are normal subgroups of MN so that MN'_r is a normal p-subgroup of MN. Since $MN = [MN'_r]N_r$, it follows now that MN'_r is the totality of p-elements in MN; and we may restate the results just obtained shortly in the form:

MN'_r is the p-component and N_r is a p-complement of the normal subgroup $MS \cap Q$ of MS where we term p-complement any subgroup of order prime to p whose index is a power of p. Note that N_r is not uniquely determined since it need not be normal in $MS \cap Q$.

We recall next that N_r is the totality of r-elements in the normal subgroup $N = Q \cap S$ of S. As a characteristic subgroup of a normal subgroup N_r is consequently a normal subgroup of S. In other words: S is part of the normalizer of N_r in MS. Consider next an element a belonging to the normalizer of N_r in MS. There exist elements m and s in M and S respectively such that $a = ms$. Since a belongs to the normalizer of N_r [by hypothesis] and since s belongs to the normalizer of N_r [as an element in S], m belongs to the normalizer of N_r. If x is an element in N_r, then the commutator $[x, m]$ belongs to N_r, since x is in the normalizer of N_r; and $[x, m]$ belongs to M, since M is a normal subgroup of G. Thus $[x, m]$ belongs to the intersection of M and N_r which is 1, since M is a p-group whereas N_r is a Pp-group. Hence $xm = mx$; and we have shown that m commutes with every element in N_r. Remembering that $N = N'_rN_r$, that $N'_r \leq Z(M < G)$, and that $Q = NZ(M < G) = N_rZ(M < G)$ it follows finally that m commutes with every element in Q. Hence m is a fixed element of every automorphism in $Q/Z(M < G)$. Since this group of auto-

morphisms of M is a normal subgroup, not 1, of the group $G/Z(M < G)$ of automorphisms of M, the group of fixed elements of $Q/Z(M < G)$ is a normal subgroup of G which is a proper part of M. It follows from the minimality of M, that this group of fixed elements equals 1. Hence $m = 1$ so that a belongs to S. Thus we have shown that

$$S \text{ is the normalizer of } N_r \text{ in } MS.$$

We are now ready to prove (c). Since the sufficiency of the condition, given in (c), is almost obvious, we assume immediately that H and K are minimal subgroups representing G in M and that $MH = MK = W$. We have shown before that the totality of p-elements in W is a characteristic subgroup W_p of W and that there exist p-complements H^* and K^* of W such that H is the normalizer of H^* in W and K is the normalizer of K^* in W. Because of the solubility of p-groups it is a special case of a Theorem of Witt-Zassenhaus that p-complements are conjugate whenever the p-Sylow subgroup is normal; see, for instance, Zassenhaus [1; p. 126, Satz 27]. Consequently H^* and K^* are conjugate in W. Hence there exists an element w in W such that $H^* = w^{-1}K^*w$. Since H and K are the normalizers in W of H^* and K^* respectively, it follows that $H = w^{-1}Kw$. Since $W = KM$, there exist elements k and x in K and M respectively such that $w = kx$. Consequently

$$H = w^{-1}Kw = x^{-1}k^{-1}Kkx = x^{-1}Kx;$$

and this completes the proof of (c).

We have already verified (a), (c) and the first part of (b); and we are going to deduce now the second part of (b) from these three properties without any further reference to the original hypotheses. Consider therefore a minimal subgroup S representing G in M. Then $M \cap S = 1$ and M is a minimal normal subgroup of MS [Lemma 1, (c)]; and S is a maximal subgroup of MS [§ 2, Lemma 1]. Consequently $\phi(MS) \leq S$. Next we deduce from the nilpotency of Frattini's subgroup and from Lemma 2, (c) that

$$\phi(MS) \leq S \cap F(MS) = S \cap Z(M < G) \cap MS = S \cap Z(M < G) = J;$$

and we deduce from Lemma 1, (b) and Lemma 2, (b) that J is a normal subgroup of MS which is part of $\phi(S)$.

Consider now some maximal subgroup T of MS. We distinguish two cases.

Case 1. $M \leq T$.

It follows from Dedekind's Law that $T = M(S \cap T)$ and that $S \cap T$ is a maximal subgroup of S. Hence $J \leq \phi(S) \leq S \cap T \leq T$.

Case 2. $M \not\leq T$.

Then $M \cap T = 1$ [§ 2, Lemma 1]; and it follows from the maximality of T that $MT = MS$. Next it is clear that

$$TZ(M < G) = TMZ(M < G) = SMZ(M < G) = SZ(M < G) = G;$$

T represents G in M. Suppose now that the subgroup X of T represents G in M. Then $G = [XM]Z(M < G)$ and $XM \leq TM = SM$ imply, by Lemma 2, (a), that $XM = SM = TM$ and that therefore $X = T$. In other words: T is a minimal subgroup representing G in M. But then we deduce from (c) the existence of an element v in M such that $T = v^{-1}Sv$. Consequently

$$J = v^{-1}Jv \leq v^{-1}Sv = T.$$

Thus we have shown that J is part of every maximal subgroup T of MS. Hence $J \leq \phi(MS)$; and this completes the proof of the equation

$$\phi(MS) = J = Z(M < G) \cap S.$$

Remark. The principal content of the preceding discussion for our future applications may be stated as follows:

(E) *If M is a minimal normal subgroup of G, and if $G/Z(M < G)$ contains a normal subgroup, not* 1, *whose order is prime to the order of M, then M is abelian and there exists a subgroup S of G such that $M \cap S = 1$ and $SZ(M < G) = G$.*

It is natural that this fundamental existence theorem (E) ceases to be true once we omit the hypothesis concerning $G/Z(M < G)$. The author is indebted to Professor Wielandt for pointing out to him the impossibility of substituting for the hypothesis concerning $G/Z(M < G)$ the weaker assumption that M be abelian.

5. Solubility. The following simple facts will be needed to obtain the connection between the results of the last section and the solubility problem.

LEMMA 1. *If the minimal normal subgroup M of G is soluble, then M is abelian and there exists a prime p such that $M^p = 1$.*

Proof. The commutator subgroup $[M, M]$ of M is a normal subgroup of G, since it is a characteristic subgroup of a normal subgroup. Furthermore $[M, M] < M$, since M is soluble. Hence $[M, M] = 1$ so that M is abelian. The existence of a prime p such that $M^p = 1$ may finally be deduced from § 2, Remark 3.

LEMMA 2. *If M is a minimal normal subgroup of G, and if $G/Z(M < G)$ contains a soluble normal subgroup different from 1, then [M is abelian and] $G/Z(M < G)$ contains a normal subgroup different from 1 whose order is prime to the order of M.*

Proof. We deduce from our hypothesis the existence of a minimal normal subgroup Q of $G/Z(M < G)$ which is soluble. It follows from Lemma 1 that Q is abelian and that there exists a prime number q such that $Q^q = 1$.

Assume now by way of contradiction that M is not part of $Z(M < G)$. Then $M \cap Z(M < G) \neq M$ and this implies $M \cap Z(M < G) = 1$, since M is a minimal normal subgroup of G. Consequently $M^* = MZ(M < G)/Z(M < G)$ is a minimal normal subgroup of $G^* = G/Z(M < G)$. We distinguish two cases.

Case 1. $M^* \cap Q \neq 1$.

Then it follows from the minimality of the normal subgroups M^* and Q of G^* that $M^* = Q$. Hence M^* is soluble. But $M^* = MZ(M < G)/Z(M < G)$ is isomorphic to M, since $M \cap Z(M < G) = 1$. Thus M is soluble. But then we infer from the minimality of M and Lemma 1 that M is abelian. Hence M is part of $Z(M < G)$ contradicting $M \cap Z(M < G) = 1$. Thus this case is impossible.

Case 2. $M^* \cap Q = 1$.

There exists a uniquely determined normal subgroup R of G which contains $Z(M < G)$ and satisfies $Q = R/Z(M < G)$. The hypothesis of case 2 is then equivalent to

$$Z(M < G) = MZ(M < G) \cap R = Z(M < G)[M \cap R],$$

as follows from Dedekind's Law. Consequently $M \cap R$ is part of $Z(M < G)$. Hence

$$M \cap R = M \cap R \cap Z(M < G) = M \cap Z(M < G) = 1.$$

Since M and R are both normal subgroups, it follows now that every element in M commutes with every element in R. Consequently $R \leq Z(M < G)$ which contradicts $Q \neq 1$. Thus we have been led again to a contradiction. Consequently M is part of $Z(M < G)$; and this implies in particular that M is abelian. We deduce from Lemma 1 the existence of a prime p such that $M^p = 1$.

Assume now by way of contradiction that $p = q$. Then Q is essentially the same as a p-group of automorphisms of the p-group M. Denote by N the group of all the fixed elements of this group Q of automorphisms of M.

Since Q is a normal subgroup of $G/Z(M < G)$, N is a normal subgroup of G. Since $Q \neq 1$, $N < M$. It follows from the minimality of M that $N = 1$. But a p-group of automorphisms of a p-group always possesses fixed elements different from 1, as may be seen from the customary arguments [like those employed when proving that the center of a p-group, not 1, is different from 1]. Thus we have have been led to a contradiction. Hence $p \neq q$ and this completes the proof.

LEMMA 3. *If $\phi(G)$ contains every normal subgroup of G except G, and if $G/\phi(G)$ is not abelian, then center and hypercenter of G are equal subgroups of $\phi(G) = F(G)$.*

Proof. G is not nilpotent, since $G/\phi(G)$ is not abelian [Wielandt's Theorem; see, for instance, § 3, Corollary 1]. Consequently $\phi(G) = F(G)$ and the center $Z(G)$ and the hypercenter $H(G)$ of G are both part of $\phi(G)$.

Suppose now that the normal subgroup N of G contains $Z(G)$ and that $N/Z(G) \leq Z[G/Z(G)]$. Then N is part of $H(G)$ so that $N \leq \phi(G)$. If n is an element in N and g an element in G, then their commutator $[n, g] = n^{-1}g^{-1}ng$ belongs to $Z(G)$. If g and h are elements in G, and if n is an element in N, then we find that

$$\begin{aligned}(gh)^{-1}n(gh) &= h^{-1}g^{-1}ngh = h^{-1}n[n, g]h = h^{-1}nh[n, g] \\ &= n[n, h][n, g] = n[n, g][n, h] = (hg)^{-1}n(hg).\end{aligned}$$

The automorphisms which are induced in N by elements in G form therefore a commutative group; and this is equivalent to saying that $G/Z(N < G)$ is abelian. Since $G/\phi(G)$ is not abelian, it is impossible that $Z(N < G)$ is part of $\phi(G)$. Hence $G = Z(N < G)$ or $N \leq Z(G)$. Consequently $N = Z(G)$; and now the equality of $Z(G)$ and $H(G)$ is easily deduced.

THEOREM 1. *The following properties of the group G are equivalent.*

(i) *G is soluble.*

(ii) *$G/\phi(G)$ is soluble.*

(iii) *If $Q \neq 1$ is a quotient group of G, then $\phi(Q) < F(Q)$.*

(iv) *If $Q \neq 1$ is a quotient group of G, then $F(Q) \neq 1$.*

(v) *If M is a minimal normal subgroup of the quotient group Q of G, and if $Z(M < Q) < Q$, then there exists a normal subgroup, not 1, of $Q/Z(M < Q)$ whose order is prime to the order of M.*

(vi) *If M is a minimal normal subgroup of the quotient group Q of*

G, then there exists a subgroup S of Q such that $Q = SZ(M < Q)$ and $1 = S \cap M$.

(vii) *If the normal subgroup $M \neq 1$ of the quotient group Q of G is part of every proper normal subgroup of Q, then $Z(M < Q) \neq 1$.*

(viii) *Every subgroup $S \neq 1$ of G has the following two properties.*

(a) $Z[F(S)] = Z[F(S) < S]$.

(b) *If $\phi(S)$ contains every normal subgroup of S, except S, and if $S/\phi(S)$ is not abelian, then $\phi(S)$ is the hypercenter of S.*

(ix) *If $S \neq 1$ is a subgroup of G, and if $\phi(S)$ contains every normal subgroup of S, except S, then $S/\phi(S)$ is abelian.*

(x) *If $S \neq 1$ is a subgroup of G, then $\phi(S) < F(S)$.*

Proof. The equivalence of properties (i) and (ii) is a fairly immediate consequence of the following facts: $\phi(G)$ is nilpotent and therefore soluble; if N is a normal subgroup of G, then solubility of N and G/N is necessary and sufficient for solubility of G.

If G is soluble and if $Q \neq 1$ is a quotient group of G, then Q too is soluble. From $Q \neq 1$ we infer $\phi(Q) < Q$. Hence $Q/\phi(Q)$ is a soluble group different from 1. But such a group contains abelian normal subgroups different from 1. It follows from § 3, Theorem 1 that

$$1 \neq F[Q/\phi(Q)] = F(Q)/\phi(Q) \text{ or } \phi(Q) < F(Q).$$

Thus (iii) is a consequence of (i); and it is obvious that (iv) is a consequence of (iii).

Assume next the validity of (iv). The group G possesses soluble normal subgroups, for instance 1; and consequently there exists a maximal soluble normal subgroup M of G. If M were different from G, then we would infer from (iv) that $F(G/M) \neq 1$. There exists one and only one normal subgroup N of G such that $M < N$ and $N/M = F(G/M)$. From the solubility of M and $F(G/M) = N/M$ we deduce now the solubility of N. This contradicts the maximality of M. Hence $M = G$ is soluble; and we have verified the equivalence of the first four conditions.

If G is soluble, and if M is a minimal normal subgroup of the quotient group Q of G such that $Z(M < Q) < Q$, then $Q/Z(M < Q)$ is soluble as a homomorphic image of the soluble group G. From $Q/Z(M < Q) \neq 1$ and Lemma 2 we deduce now the existence of a normal subgroup, not 1, of $Q/Z(M < Q)$ whose order is prime to the order of M. Thus (v) is a consequence of (i).

Assume next the validity of (v) and consider a minimal normal subgroup M of the quotient group Q of G. If $Z(M < Q) = Q$, then $S = 1$ meets the requirements of (vi). If on the other hand $Z(M < Q) < Q$, then it follows from (v) that $Q/Z(M < Q)$ contains a normal subgroup, not 1, whose order is prime to the order of M; and we deduce from § 4, Proposition 2 [or § 4, (E)] the existence of a subgroup S of Q such that $Q = SZ(M < Q)$ and $1 = M \cap S$. Hence (vi) is a consequence of (v).

Assume next the validity of (vi) and consider a normal subgroup $M \neq 1$ of the quotient group Q of G such that every proper normal subgroup of Q contains M. Then M is a minimal normal subgroup of Q; and we deduce from (vi) the existence of a subgroup S of Q such that $Q = SZ(M < Q)$ and $1 = M \cap S$. If $Z(M < Q)$ were equal to 1, then S would equal Q and this would imply $M = 1$, an impossibility. Hence $Z(M < Q) \neq 1$ so that (vii) is a consequence of (vi).

Assume next the validity of (vii) and assume by way of contradiction that G is not soluble. Then G possesses normal subgroups V such that G/V is not soluble, for instance $V = 1$; and among these normal subgroups there exists a maximal one W. Then $Q = G/W$ is not soluble; but if $U \neq 1$ is a normal subgroup of Q, then Q/U is soluble. Clearly $Q \neq 1$ and consequently there exists a minimal normal subgroup M of Q. From $M \neq 1$ we infer the solubility of Q/M. Now we distinguish two cases.

Case 1. There exists a proper normal subgroup N of Q which does not contain M.

From $N \neq 1$ we infer the solubility of Q/N. Since the minimal normal subgroup M of Q is not part of N, we have $M \cap N = 1$. Hence M is isomorphic to the subgroup NM/N of the soluble group Q/N. Consequently the minimal normal subgroup M is soluble; and it follows from Lemma 1 that M is abelian.

Case 2. M is part of every proper normal subgroup of Q.

Then we apply (vii) and find that $Z(M < Q) \neq 1$. Hence M is part of $Z(M < Q)$; and we see again that M is abelian.

Thus we have shown that M and Q/M are both soluble; and this proves the solubility of Q, providing us with the desired contradiction. Hence G is soluble; and we have verified the equivalence of the first seven conditions.

Assume again the solubility of G and consider a subgroup $S \neq 1$ of G. Then S too is soluble. This implies in particular the solubility of $Z[F(S) < S]/Z[F(S)]$; and it follows from § 3, Proposition 2, that

$Z[F(S)] = Z[F(S) < S]$. Assume next that $\phi(S)$ contains every normal subgroup of S, except S. Since S is soluble, it follows from (iii) that $\phi(S) < F(S)$. Hence $S = F(S)$ is nilpotent and this implies commutativity of $S/\phi(S)$; see, for instance § 3, Corollary 1. Thus (viii) is a consequence of (i).

Assume next the validity of (viii) and consider a subgroup $S \neq 1$ of G such that $\phi(S)$ contains every normal subgroup of S except S. If $S/\phi(S)$ were not abelian, then $\phi(S)$ would, by (viii. b), be the hypercenter of S. But now we could infer from Lemma 3 that $\phi(S)$ is the center of S. Since $S/\phi(S)$ is not abelian, it is furthermore impossible that $S = F(S)$ [§ 3, Theorem 1]. Hence $F(S) = \phi(S)$ is the center of S so that

$$Z[F(S)] \leq F(S) = \phi(S) < S = Z[F(S) < S],$$

contradicting (viii. a). Hence $S/\phi(S)$ is abelian; and we have shown that (ix) is a consequence of (viii).

Assume next the validity of (ix). If G were not soluble, then there would exist a minimal subgroup S of G such that S is not soluble. Thus S itself is not soluble, but every proper subgroup of S is soluble. Consider now a normal subgroup N of S which is different from S and not part of $\phi(S)$. Then $N < S$ implies the solubility of N. Since N is not part of $\phi(S)$, there exists a maximal subgroup T of S which does not contain N. Clearly T is soluble and $S = NT$. But then $S/N \simeq T/(T \cap N)$ is soluble as a homomorphic image of the soluble group T. The solubility of N and S/N implies the solubility of S which is impossible. Thus we have shown that every normal subgroup of S with the exception of S itself is part of $\phi(S)$; and now it follows from (ix) that $S/\phi(S)$ is abelian. But $\phi(S)$ is nilpotent and consequently soluble proving again the solubility of S. Our hypothesis that G is not soluble has led us to a contradiction which proves the equivalence of properties (i) to (ix).

If G is soluble, then every subgroup $S \neq 1$ of G is soluble too; and $\phi(S) < F(S)$ may be derived from (iii). Thus (x) is a consequence of (i). Assume conversely the validity of (x) and consider a subgroup $S \neq 1$ of G such that every normal subgroup, not S, of S is part of $\phi(S)$. It follows from (x) that $\phi(S) < F(S)$ and that therefore $F(S) = S$. The commutativity of $S/\phi(S) = F(S)/\phi(S)$ is now a consequence of § 3, Theorem 1. Hence (ix) is a consequence of (x); and this completes the proof of the equivalence of our ten properties.

Remark. The author is indebted to Professor Wielandt for pointing out to him a class of groups which shows the impossibility of essentially weakening

condition (viii). This is the more interesting, since one may deduce from (ix) the absence of subgroups S of soluble groups G with the following properties: $\phi(S)$ contains every proper normal subgroup of S and $S/\phi(S)$ is not abelian. Hence (viii. b) is somewhat vacuously satisfied by soluble groups.

6. *n*-soluble groups. For a convenient enunciation of our subsequent results we need the following concepts which have been introduced elsewhere; see Baer [3, 4]. The element z in the group G is termed *n-central,* for n an integer, if

$$(zg)^n = z^n g^n \text{ and } (gz)^n = g^n z^n \text{ for every } g \text{ in } G;$$

and the totality of n-central elements in G is the *n-center* $Z(G;n)$ of G. The n-center is a characteristic subgroup and $Z(G;n) = Z(G;1-n)$. Similarly we term the group G *n-abelian,* if

$$(xy)^n = x^n y^n \text{ for every } x \text{ and } y \text{ in } G.$$

It has been shown in Baer [4] that the study of n-abelian groups may effectively be reduced to the study of n-abelian n-groups.

PROPOSITION 1. *If G is an n-group and $\phi(G) = 1$, then the following properties of G are equivalent.*

(i) *G is n-abelian.*

(ii) *Every minimal normal subgroup of G belongs to the n-center of G.*

(iii) $G^n = 1$.

Proof. It is trivial that (i) implies (ii) and that (iii) implies (i). Thus we need show only that (iii) is a consequence of (ii). Assume consequently the validity of (ii); and assume by way of contradiction that $G^n \neq 1$. Then there exists a minimal normal subgroup M of G which is part of G^n; and we deduce $M \leq Z(G;n)$ from (ii). Since G is an n-group, so is M; and this implies that every element in M is the $(1-n)$-th power of an element in M. But the $(1-n)$-th powers of elements in the n-center commute with every n-th power; see Baer [3; (2. 4, b)]. Consequently G^n is part of the centralizer of M in G so that

$$M \leq G^n \leq Z(M < G).$$

Hence M is abelian; and we deduce from the minimality of M the existence of a prime p such that $M^p = 1$ [§ 5, Lemma 1]. Since M is an n-group, p is a divisor of n. From $\phi(G) = 1$ we deduce the existence of a maximal

subgroup S of G which does not contain the abelian minimal normal subgroup M of G. It follows from § 2, Lemma 1 that S is a complement of M in G. Consider now an element g in G. Then there exist uniquely determined elements m and s in M and S respectively such that $g = ms$. Since m, as an element in $M \leq Z(G; n)$, is n-central, we find that $g^n = (ms)^n = m^n s^n = s^n$. [Remember that $M^p = 1$ and that p is a divisor of n.] Hence $G^n \leq S$ so that $1 < M = M \cap G^n \leq M \cap S = 1$, a contradiction which proves $G^n = 1$. This completes the proof.

COROLLARY 1. *The group G without proper characteristic subgroups is n-abelian if, and only if, G is abelian or $G^n = 1$ or $G^{1-n} = 1$.*

Proof. The sufficiency of our condition is almost obvious. Assume therefore that G is n-abelian. Then it is known that G is the direct product of an n-abelian n-group G_n, a $(1-n)$-abelian $(1-n)$-group G_{1-n} and of an abelian $Pn(1-n)$-group A; see Baer [4]. It is clear that G_n, G_{1-n} and A are characteristic subgroups and that therefore only one of them may be different from 1. It follows furthermore from $\phi(G) < G$ and the absence of proper characteristic subgroups that $\phi(G) = 1$. Now we deduce $G^n = 1$ from Proposition 1, in case $G = G_n$; and we deduce $G^{1-n} = 1$ from Proposition 1, in case $G = G_{1-n}$. This completes the proof.

A group G has been termed *n-soluble,* if every quotient group $Q \neq 1$ of G possesses an n-abelian minimal normal subgroup; see Baer [4]. This is clearly equivalent to the requirement that every principal factor of G is n-abelian. [For the definition of principal factors see Zassenhaus [1; p. 53].] Thus we obtain the following consequence of Corollary 1.

PROPOSITION 2. *The group G is n-soluble if, and only if, every principal factor C of G satisfies one of the following three conditions:*

$$C \text{ is abelian or } C^n = 1 \text{ or } C^{1-n} = 1.$$

7. n-nilpotent n-groups. A group G has been termed *n-nilpotent,* if the n-center of each of its quotient groups, not 1, is different from 1; see Baer [4]. It is not difficult to see that n-nilpotent groups have the following stronger property:

(N-n) *If M is a minimal normal subgroup of the quotient group Q of G, then M is part of the n-center of Q.*

It has been shown in Baer [4] that the study of n-nilpotent groups may

effectively be reduced to the study of n-nilpotent n-groups and thus these will be the object of our present investigation.

PROPOSITION 1. *If the n-group G is n-nilpotent, then $G^n \leq \phi(G)$.*

Proof. Let $G^* = G/\phi(G)$. Then G^* is an n-nilpotent n-group such that $\phi(G^*) = 1$. Hence it follows from (N-n) and § 6, Proposition 1 that $G^{*n} = 1$; and this is equivalent to $G^n \leq \phi(G)$.

Remark 1. It is not difficult to prove that n-nilpotency of a random group implies n-commutativity of $G/\phi(G)$.

PROPOSITION 2. *The soluble n-group G is n-nilpotent if, and only if,*

$$S^n \leq \phi(S) \text{ for every subgroup } S \text{ of } G.$$

Proof. Since every subgroup of an n-nilpotent n-group is itself an n-nilpotent n-group, we may derive the necessity of our condition from Proposition 1.

Assume conversely the validity of our condition. Consider a quotient group $Q = G/N \neq 1$ of G and a minimal normal subgroup $M/N = M^*$ of Q. From the solubility of G, and hence of Q, we deduce, by § 5, Theorem 1, (vi) the existence of a subgroup S^* of Q such that $1 = M^* \cap S^*$ and $Q = S^*Z(M^* < Q)$. There exist uniquely determined subgroups S and V of G which both contain N and satisfy $S^* = S/N$ and $Z(M^* < Q) = V/N$. It is clear that V is a normal subgroup of G and that $N = M \cap S$ and $G = VS$.

Next we let $T = MS$ and $T^* = T/N = M^*S^*$. Since the elements in S^* induce in M^* all the automorphisms that are induced in M^* by elements in G^*, M^* is likewise a minimal normal subgroup of T^*. Since S^* is a complement of M^* in T^*, it follows from § 2, Lemma 1 that S^* is a maximal subgroup of T^*. Hence S is a maximal subgroup of T; and it follows now from our condition that $T^n \leq \phi(T) \leq S$, and that consequently $T^{*n} \leq S^*$. Consider now an element t^* in T^* and an element m^* in M^*. Then $(t^*m^*)^n(t^{*n}m^{*n})^{-1}$ and $(m^*t^*)^n(m^{*n}t^{*n})^{-1}$ belong to T^{*n} and therefore to S^*; and they also belong to M^*, since M^* is a normal subgroup of T^*. These elements belong therefore to the intersection 1 of M^* and S^*; and this shows that

$$(t^*m^*)^n = t^{*n}m^{*n} \text{ and } (m^*t^*)^n = m^{*n}t^{*n} \text{ for } m^* \text{ in } M^* \text{ and } t^* \text{ in } T^*.$$

Consider now an element m^* in M^* and an element q in Q. Then there exists an element s^* in S^* such that $q \equiv s^*$ modulo $Z(M^* < Q)$. The

elements q and s^* induce therefore the same automorphism in M^* and this implies in particular

$$q^{-i}m^*q^i = s^{*-i}m^*s^{*i} \text{ for every integer } i.$$

Application of this and the preceding formula gives us now:

$$\begin{aligned}(m^*q)^n &= m^*(qm^*q^{-1})(q^2m^*q^{-2}) \cdots (q^{n-1}m^*q^{1-n})q^n \\ &= m^*(s^*m^*s^{*-1})(s^{*2}m^*s^{*-2}) \cdots (s^{*n-1}m^*s^{*1-n})q^n \\ &= (m^*s^*)^n s^{*-n}q^n = m^{*n}q^n;\end{aligned}$$

and $(qm^*)^n = q^n m^{*n}$ is shown likewise. Thus we have finally verified that M^* is part of the n-center of Q. We have therefore proven the validity of (N-n) and consequently the n-nilpotency of G.

Remark 2. The author has not been able to decide whether the solubility hypothesis may be omitted from Proposition 2 nor could he decide whether the weaker hypothesis $G^n \leqq \phi(G)$ is sufficient.

University of Illinois,
Urbana, Illinois.

BIBLIOGRAPHY.

Reinhold Baer [1], "Nilpotent groups and their generalizations," *Transactions of the American Mathematical Society*, vol. 47 (1940), pp. 393-434.

——— [2], "Extension types of abelian groups," *American Journal of Mathematics*, vol. 71 (1949), pp. 461-490.

——— [3], "Endlichkeitskriterien für Kommutatorgruppen," *Mathematische Annalen*, vol. 124 (1952), pp. 161-170.

——— [4], "Factorization of n-soluble and n-nilpotent groups," *Proceedings of the American Mathematical Society*, vol. 4 (1953), pp. 15-26.

——— [5], "Group elements of prime power index," *Transactions of the American Mathematical Society*, to appear shortly.

H. Fitting [1], "Beiträge zur Theorie der Gruppen endlicher Ordnung," *Jahresberichte der Deutschen Mathematiker Vereinigung*, vol. 48 (1938), pp. 77-141.

W. Gaschütz [1], "Über die ϕ-Untergruppe endlicher Gruppen," *Mathematische Zeitschrift*, vol. 58 (1953), pp. 160-170.

H. Zassenhaus [1], "*Lehrbuch der Gruppentheorie I*," Leipzig-Berlin (1937).

ON THE STRUCTURE OF UNITARY GROUPS (II).*

By JEAN DIEUDONNÉ.

1. This paper adds some miscellaneous results on unitary groups to those which have been proved in [4]. In sections 2 to 6, I show how the study of unitary groups over a field of characteristic 2 can always be reduced to the case, considered in [4], in which the hermitian form is "trace-valued." In section 8 to 12 I prove that, with a single exception, quasi-symmetries generate the unitary group, and deduce from that fact certain information on the determinant of a unitary transformation.

2. The terminology and notations are those of [4]. When K is a sfield of characteristic 2 and f an arbitrary nondegenerate hermitian form over the n-dimensional space E, we are going to see that the structure of the group $U_n(K, f)$ can essentially be reduced to that of another unitary group $U_m(K, f_1)$, where f_1 is a "trace-valued" form, that is, such that every value $f_1(y, y)$ in K can be written $\xi + \xi^J$. We observe here that the case in which K is commutative and J the identity is included in what follows, and gives back the treatment of the groups leaving invariant a symmetric form over a field of characteristic 2 ([2], p. 60), of which the following is obviously a generalization.

Let V be the subset of E consisting of vectors x such that $f(x, x)$ has the form $\xi + \xi^J$; owing to the formulas

$$f(x + y, x + y) = f(x, x) + f(y, y) + f(x, y) + (f(x, y))^J$$

and

$$f(x\lambda, x\lambda) = \lambda^J f(x, x)\lambda$$

V is a vector subspace of E. Let V^* be the subspace orthogonal to V, $V_1 = V \cap V^*$, V_2 a subspace of V supplementary to V_1; let q be the dimension of V_1, m that of V_2. V^* has, then, dimension $n - (m + q)$; let V_3 be a subspace of V^* supplementary to V_1, of dimension $n - m - 2q$. V_2 and V_3 are non-isotropic, and are orthogonal to each other, therefore $V_2 + V_3$ is non-isotropic; so is therefore $(V_2 + V_3)^*$, which has dimension $2q$ and is supplementary to $V_2 + V_3$. For future purposes, we prove the following lemma:

* Received January 29, 1953.

LEMMA 1. *There exists a basis* e_i $(1 \leq i \leq 2q)$ *of* $(V_2 + V_3)^* = V_2^* \cap V_3^*$ *such that the vectors* $e_1, e_2, \cdots, e_q$ *form a basis of* V_1 *and that* $f(e_i, e_{q+j}) = 0$ *for* $i \neq j$, $f(e_i, e_{q+i}) = 1$ *for* $1 \leq i \leq q$.

It is clear that V_1 is contained in $V_2^* \cap V_3^*$. Let $e_1 \neq 0$ be an arbitrary vector in V_1; e_1 cannot be orthogonal to $V_2^* \cap V_3^*$, for as it is already orthogonal to $V_2 + V_3$, it would be orthogonal to E, which is contrary to the assumption that f is nondegenerate. There is therefore a vector e_{q+1} in $V_2^* \cap V_3^*$ such that $f(e_1, e_{q+1}) \neq 0$, and as e_1 is orthogonal to V_1, e_{q+1} is not in V_1; by multiplication of e_{q+1} by a scalar, we can suppose that $f(e_1, e_{q+1}) = 1$. As $f(e_1, e_1) = 0$, it is readily verified that the restriction of f to the plane P generated by e_1 and e_{q+1} is nondegenerate. Therefore the $(n-2)$-dimensional subspace P^* orthogonal to P is supplementary to P; moreover, P^* contains $V_2 + V_3$, and the hyperplane H generated by P^* and e_1 is orthogonal to e_1 and contains V_1; it follows from this that the intersection $P^* \cap V_1$ is $(q-1)$-dimensional. We can then prove the lemma by induction on q, since the restriction of f to P^* is nondegenerate, and the subspace of P^* where $f(x, x)$ is "trace-valued" consists of $V \cap P^* = V_2 + (V_1 \cap P^*)$; there exists therefore a basis $e_2, \cdots, e_q, e_{q+2}, \cdots, e_{2q}$ of $P^* \cap (V_2^* \cap V_3^*)$ such that $e_2, \cdots, e_q$ form a basis for $P^* \cap V_1$, and $f(e_i, e_{q+j}) = \delta_{ij}$ for $i \geq 2$ and $j \geq 2$; it is then clear that e_i $(1 \leq i \leq 2q)$ verify the conditions of the lemma.

We shall designate by V_4 the subspace generated by $e_{q+1}, \cdots, e_{2q}$; E is therefore the direct sum of the 4 subspaces V_1, V_2, V_3, V_4.

3. Let u be an arbitrary transformation in the unitary group $U_n(K, f)$; for every $x \in V_3 + V_4$, we can write $u(x) = v(x) + w(x)$, where $v(x) \in V_3 + V_4$ and $w(x) \in V = V_1 + V_2$; the relation $f(u(x), u(x)) = f(x, x)$ can be written

$$f(v(x), v(x)) + f(w(x), w(x)) + f(v(x), w(x)) + f(v(x), w(x))^J = f(x, x)$$

hence

$$f(v(x) - x, v(x) - x) \\ = f(v(x), w(x) - x) + (f(v(x), w(x) - x))^J + f(w(x), w(x)).$$

But as $w(x) \in V$, $f(w(x), w(x))$ can be written $\lambda + \lambda^J$ by assumption, hence the same conclusion is true of $f(v(x) - x, v(x) - x)$, from which it follows, by definition, that $v(x) - x \in V$; but $v(x) - x$ is by assumption in $V_3 + V_4$, which proves that $v(x) = x$ in $V_3 + V_4$, hence $u(x) = x + w(x)$.

We next remark that u leaves V invariant (globally) by definition,

hence it also leaves V^* globally invariant, and the same is true therefore of $V_1 = V \cap V^*$. Let $x \varepsilon V_1$, $y \varepsilon V_4$; the relation $f(u(x), u(y)) = f(x, y)$ yields

$$f(u(x), y + w(y)) = f(x, y)$$

and as $w(y) \varepsilon V$ is orthogonal to $u(x) \varepsilon V_1$, this is equivalent to $f(u(x) - x, y) = 0$. In other words, $u(x) - x$ is a vector in V_1 which is orthogonal to V_4; but such a vector is therefore orthogonal to E (since V_1 is orthogonal to $V_1 + V_2 + V_3$), hence 0; in other words, $u(x) = x$ for every $x \varepsilon V_1$.

Finally, for $x \varepsilon V_3$, we must have $u(x) = x + w(x) \varepsilon V^*$, hence $w(x) \varepsilon V^*$; but as $w(x) \varepsilon V$, we have $w(x) \varepsilon V_1$; for every $y \varepsilon V_4$, we write $f(u(x), u(y)) = f(x, y)$, or equivalently

$$f(x + w(x), y + w(y)) = f(x, y).$$

As x is orthogonal to $w(y) \varepsilon V$, and $w(x) \varepsilon V_1$ is also orthogonal to $w(y)$, this relation reduces to $f(w(x), y) = 0$ for every $y \varepsilon V_4$; the same argument as above proves then that $w(x) = 0$.

Summing up, we see that $u(x) = x$ in $V^* = V_1 + V_3$, and that in V_4 we may write $u(x) = x + w_1(x) + w_2(x)$, with $w_1(x) \varepsilon V_1$ and $w_2(x) \varepsilon V_2$.

4. As u leaves invariant (globally) $V = V_1 + V_2$, for $x \varepsilon V_2$ we can write $u(x) = u_0(x) + v_1(x)$, where $u_0(x) \varepsilon V_2$ and $v_1(x) \varepsilon V_1$; if $y \varepsilon V_2$, the relation $f(u(x), u(y)) = f(x, y)$ yields (owing to the fact that V_1 is orthogonal to V)

$$f(u_0(x), u_0(y)) = f(x, y).$$

This shows that u_0 is a mapping of V_2 into itself which belongs to the unitary group $U_m(K, f_1)$, where f_1 is the restriction of f to V_2; due to the definition of V_2, f_1 is *nondegenerate* and *trace-valued*, and therefore the study of the group $U_m(K, f_1)$ can be made with the methods developed in [4].

In order to write that u satisfies the identity $f(u(x), u(y)) = f(x, y)$ for all pairs of vectors x, y in E, it remains only to write this identity when $x \varepsilon V_2$ and $y \varepsilon V_4$, or when both x and y are in V_4; this yields the two relations

(1) $\quad f(v_1(x), y) + f(u_0(x), w_2(y)) = 0$ for $x \varepsilon V_2$ and $y \varepsilon V_4$;

(2) $\quad f(x, w_1(y)) + f(w_1(x), y) + f(w_2(x), w_2(y)) = 0$ for $x \varepsilon V_4$, $y \varepsilon V_4$.

We want to prove the following

LEMMA 2. *For an arbitrary unitary mapping $u_0 \varepsilon U_m(K, f_1)$, and an arbitrary linear mapping v_1 from V_2 into V_1, there exists at least one unitary mapping $u \varepsilon U_n(K, f)$ whose restriction to V_2 is $u_0 + v_1$.*

Let us prove first that equation (1) determines entirely the linear mapping w_2 of V_4 into V_2. Indeed, for any given $y \varepsilon V_4$, the mapping $x \rightarrow f(y, v_1(x))$ is a linear form over the vector space V_2; as the restriction of f to V_2 is a nondegenerate bilinear hermitian form, there is one and only one vector $z \varepsilon V_2$ such that $f(y, v_1(x)) = f(z, x)$ for all $x \varepsilon V_2$, and it is clear that z is a linear function of $y \varepsilon V_4$; we can therefore write identically $f(y, v_1(x)) = f(h(y), x)$, where h is a linear mapping of V_4 into V_2. On the other hand, we have $f(h(y), x) = f(u_0(h(y)), u_0(x))$ since $h(y) \varepsilon V_2$; this shows that $w_2(y) = u_0(h(y))$, and proves our assertion.

We have now to determine w_1 from equation (2); it will be enough to verify the equations

$$(3) \qquad f(e_{q+i}, w_1(e_{q+j})) + f(w_1(e_{q+i}), e_{q+j}) = f(w_2(e_{q+i}), w_2(e_{q+j}))$$

for all values of $i \leq q$ and $j \leq q$. Let $w_1(e_{q+i}) = \sum_{j=1}^{q} e_j \alpha_{ij}$; equations (3) become then

$$(4) \qquad \alpha_{ji} + \alpha_{ij}{}^J = \rho_{ij}$$

where $\rho_{ij} = f(w_2(e_{q+i}), w_2(e_{q+j}))$. Now observe that as w_2 maps V_4 into V_2, ρ_{ii} can be written as a "trace" $\gamma_i + \gamma_i{}^J$; on the other hand $\rho_{ji} = \rho_{ij}{}^J$. We can therefore solve equations (4) by taking for instance $\alpha_{ii} = \gamma_i$ and $\alpha_{ij} = \rho_{ij}{}^J$ for $i < j$, $\alpha_{ij} = 0$ for $i > j$. Lemma 2 is thus completely proved.

5. For every $u \varepsilon U_n(K, f)$, let u_V be the restriction of u to the subspace V; the mapping $u \rightarrow u_V$ is a homomorphism of U_n onto a group G_V of linear mappings of V into itself, leaving every element of V_1 invariant. We verify immediately that in G_V the subgroup G_0 consisting of the u_V such that $u_0(x) = x$ in V_2 is an abelian normal subgroup of G_V, isomorphic to the additive group of all linear mappings of V_2 into V_1, or equivalently to the additive abelian group K^{mq}. G_0 is the kernel of the homomorphism $u_V \rightarrow u_0$, and this homomorphism maps G_V on the group $U_m(K, f_1)$, since it follows from Lemma 2 that u_0 may be chosen arbitrarily in that last group; therefore G_V/G_0 is isomorphic to $U_m(K, f_1)$. Finally, the kernel of $u \rightarrow u_V$ is the normal subgroup Γ of U_n consisting of the transformations u such that $v_1(x) = 0$ and $u_0(x) = x$ for $x \varepsilon V_2$; equation (1) then yields $f(x, w_2(y)) = 0$ for $x \varepsilon V_2$ and $y \varepsilon V_4$; as $w_2(y) \varepsilon V_2$ and the restriction of f to V_2 is nondegenerate, $w_2(y) = 0$ for all $y \varepsilon V_4$. Equation (2) then reduces to

$$f(x, w_1(y)) + f(w_1(x), y) = 0$$

for $x \in V_4$ and $y \in V_4$; and the same argument as in the proof of Lemma 2 shows that this equation is equivalent to the relations

$$\alpha_{ji} + \alpha_{ij}{}^J = 0 \tag{5}$$

for the elements of the matrix of w_1. This determines the elements α_{ji} for $i < j$ when the elements α_{ij} for $i < j$ are taken arbitrarily in K; on the other hand, relation (5) for $i = j$ shows that α_{ii} must belong to the set S of symmetric elements of K. As it is readily seen that the group Γ is isomorphic to the additive group of the matrices of the mappings w_1, this group is therefore isomorphic to $S^q \times K^{q(q-1)/2}$.

Summing up the preceding results, we finally have

THEOREM 1. *The group $U_n(K, f)$ has a composition series $U_n \supset \Gamma_0 \supset \Gamma$ such that $U_n(K, f)/\Gamma_0$ is isomorphic to $U_m(K, f_1)$, where $m \leq n$ and f_1 is a nondegenerate, trace-valued hermitian form; Γ_0/Γ is an abelian group isomorphic to the additive group K^{mq} and Γ is an abelian group isomorphic to the additive group $S^q \times K^{q(q-1)/2}$ $(2q \leq n - m)$.*

6. In addition to the condition $m + 2q \leq n$, the numbers m and q may have to satisfy additional restrictions due to the nature of the sfield K. For instance, suppose K is the reflexive sfield of rank 4 over the field $Z = \mathfrak{F}_2(X)$ (field of rational functions of one indeterminate over the prime field $\mathfrak{F}_2$ of 2 elements), with basis 1, ω, θ and $\theta\omega$ such that $\omega^2 = \omega + 1$, $\theta^2 = X$ and $\theta\omega\theta^{-1} = \omega + 1$, ([2], p. 73); 1, θ and $\theta\omega$ constitute a basis for the set of symmetric elements, and the "traces" $\xi + \xi^J$ are the elements of Z. Now, as $f(x, x)$ must not be equal to a trace in $V_3 + V_4$, except for $x = 0$, and in particular $f(x, x) \neq 0$ for $x \neq 0$ in $V_3 + V_4$, a classical argument shows that there exists in $V_3 + V_4$ an orthogonal basis (c_i) $(1 \leq i \leq n - m - q)$, that is, such that $f(c_i, c_j) = 0$ for $i \neq j$. In order that $f(x, x)$ be distinct from a trace for $x \neq 0$ in $V_3 + V_4$, it is then necessary and sufficient that $f(c_i, c_i) = a_i\theta + b_i\theta\omega$, with a_i and b_i in Z and not both 0, and that the equations $\sum_{i=1}^{n-m-q} a_i N(\xi_i) = 0$ and $\sum_{i=1}^{n-m-q} b_i N(\xi_i) = 0$ have no solution other than $\xi_i = 0$ in the sfield K. But Z is a quadratic inseparable extension of its subfield $Z^2 = \mathfrak{F}_2(X^2)$, so that we may write $a_i = p_i^2 + Xq_i^2$, $b_i = r_i^2 + Xs_i^2$, with p_i, q_i, r_i, s_i in Z. Now if we take ξ_i in Z, $N(\xi_i) = \xi_i^2$, and the equations $\sum_i a_i N(\xi_i) = 0$ and $\sum_i b_i N(\xi_i) = 0$ are equivalent to the system of 4 linear equations in the ξ_i, $\sum_i p_i\xi_i = 0$, $\sum_i q_i\xi_i = 0$, $\sum_i r_i\xi_i = 0$, $\sum_i s_i\xi_i = 0$. Such a system has a nontrivial solution in Z as soon as $n - m - q > 4$, hence we

get the inequality $m + q \geq n - 4$, which, coupled with $m + 2q \leq n$, yields in particular $q \leq 4$ and $m \geq n - 8$.

7. Theorem 1 makes it possible to study the centralizer of an *involution* s in a group $U_n(K, f)$, where K has characteristic 2 and f is a nondegenerate trace-valued hermitian form. Such a transformation has the form $x \rightarrow x + t(x)$, where $x \rightarrow t(x)$ is a linear mapping of E onto a totally isotropic subspace U of dimension $p \leq n/2$, whose orthogonal subspace $U^* \supset U$ is equal to the kernel $t^{-1}(0)$ (invariant elements of s). Let V be a non-isotropic $(n - 2p)$-dimensional subspace, supplementary to U in U^* and let W be a totally isotropic p-dimensional subspace supplementary to U in the subspace V^* orthogonal to V.

The determination of the centralizer Γ of s follows the same method as in ([3], section 25, pp. 35-37), and we refer therefore the reader to that paper for detailed proofs. We first consider the normal subgroup Γ_1 of Γ consisting of the transformations u leaving invariant every element in U, and the normal subgroup $\Gamma_2 \subset \Gamma_1$ of Γ, consisting of transformations leaving invariant every element of U^*; it is easily verified that Γ_2 is an abelian group. The factor group Γ_1/Γ_2 is isomorphic to the group Γ_1' consisting of the restrictions to U^* of transformations belonging to Γ_1. Within this last group we consider the normal subgroup Γ_2' consisting of the elements which leave invariant every class mod. U in the space U^*, and Γ_2' is again abelian.

On the other hand, for any element $v \in \Gamma_1'$ and every $z \in V$ we may write $v(z) = v'(z) + v''(z)$, where $v'(z) \in V$ and $v''(z) \in U$; $v \rightarrow v'$ is a homomorphism of Γ_1' onto a group Γ'' of transformations of V, which is seen to be identical with $U_{n-2p}(K, f_1)$, where f_1 is the restriction of f to V; moreover, the kernel of $v \rightarrow v'$ is Γ_2', so that Γ_1'/Γ_2' is isomorphic to $U_{n-2p}(K, f_1)$.

Finally, for every $u \in \Gamma$ and every $x \in W$, let $w(x)$ be the component of $u(x)$ on W (in the decomposition of E as a direct sum of U^* and W); let $\bar{u}(x) = w(x)$ for $x \in W$, $\bar{u}(x) = u(x)$ for $x \in U$ and $\bar{u}(x) = x$ for $x \in V$. It can be verified that $\bar{u} \in \Gamma$, is in the same class as u mod. Γ_1, and that $u \rightarrow \bar{u}$ is a homomorphism of Γ onto a subgroup $\bar{\Gamma}$ of Γ, isomorphic to Γ/Γ_1, consisting of transformations leaving invariant every element of V (and which can therefore be considered as unitary transformations in the nonisotropic subspace V^* of E), and leaving invariant (globally) U and W. If we take a basis of V^* consisting of p vectors e_i in U and p vectors e_{p+i} in W, such that $f(e_i, e_{p+j}) = \delta_{ij}$ (see [2], p. 6 and [4], p. 369), the matrix of $\bar{u}$ has then the form $\begin{pmatrix} A & 0 \\ 0 & \check{A}^J \end{pmatrix}$ where A is a square matrix of order p and $\check{A}$ its

contragredient. For that same basis, the matrix of s is $\begin{pmatrix} I & S \\ 0 & I \end{pmatrix}$, where S is a hermitian matrix of order p (i. e., such that ${}^tS = S^J$) which is non-degenerate, but can correspond to a non-trace-valued form f_2. The condition of permutability for $\bar{u}$ and s then reduces to $AS({}^tA^J) = S$, and this proves of course that the group $\bar{\Gamma}$ is isomorphic to $U_p(K, f_2)$. Theorem 1 now shows that this group has a composition series in which the factor groups are abelian, with the exception of one group isomorphic to $U_m(K, f_3)$, where $m \leq p$ and f_3 is a trace-valued form.

Summing up, we conclude that *the group* Γ *has a composition series in which all factor groups are abelian, with the exception of two groups isomorphic to* $U_{n-2p}(K, f_1)$ *and* $U_m(K, f_3)$, *where* $m \leq p$ (possibly $m = 0$, in which case $U_m(K, f_3)$ is understood to be reduced to the identity), *and where* f_1 *and* f_3 *are trace-valued forms.*

8. In this section, K is an arbitrary sfield with an involution J, and f a *hermitian* nondegenerate form over an n-dimensional right vector space E over K; f is supposed to be trace-valued when K has characteristic 2, and we exclude the case of the symplectic groups over fields of characteristic 2; in other words, we always suppose there are vectors $x \in E$ such that $f(x, x) \neq 0$. For such a nonisotropic vector a, a *quasi-symmetry* of hyperplane H orthogonal to a is a unitary transformation of $U_n(K, f)$ leaving invariant every element of H; there are always quasi-symmetries of hyperplane H not reduced to the identity ([4], p. 370).

THEOREM 2. *The group* $U_n(K, f)$ *is generated by quasi-symmetries except when* $n = 2$, $K = \mathfrak{F}_4$ *and* $J \neq 1$.

Of course, for orthogonal groups over fields of characteristic $\neq 2$, this reduces to the well-known theorem that the group is generated by symmetries, ([2], p. 20, prop. 8).

The theorem being obvious for $n = 1$, we use induction on n. Let $u \in U_n(K, f)$, and suppose we can find a nonisotropic vector x and a scalar $\alpha \in K$ with the following properties: 1° $f(x\alpha, x\alpha) = \alpha^J f(x, x)\alpha = f(x, x)$; 2° $u(x) - x\alpha$ is not isotropic. Then, if H is the hyperplane orthogonal to $u(x) - x\alpha$, one can write $x\alpha = y + z$, where $y \in H$ and z is orthogonal to H, and similarly $u(x) = y + z'$, where z' is orthogonal to H; as $f(u(x), u(x)) = f(x, x) = f(x\alpha, x\alpha)$, we have also $f(z, z) = f(z', z')$, hence, as $z' = z\gamma$, $\gamma^J f(z, z)\gamma = f(z, z)$. It follows that the quasi-symmetry s of hyperplane H, transforming z into $z\gamma = z'$, will transform $x\alpha$ into $u(x)$. A similar argument

proves that there is a quasi-symmetry s', whose hyperplane is orthogonal to x, and which transforms x into $x\alpha$; then $(ss')^{-1}u$ will leave x invariant, and can be considered as a unitary transformation operating in the hyperplane orthogonal to x. As such, it is a product of quasi-symmetries by the induction hypothesis, and so therefore is u.

If the characteristic of K is not 2, then for every nonisotropic vector x, one of the scalars 1 or -1 will satisfy both conditions 1° and 2°, for $u(x)-x$ and $u(x)+x$ cannot both be isotropic, as this would imply $f(x,x)+f(u(x),u(x))=0$, or $2f(x,x)=0$, contrary to assumption. We can therefore restrict our further arguments to the case in which *K has characteristic* 2 and $J\neq 1$ (otherwise, f, being trace-valued, would be an alternate form), and for every *nonisotropic vector* x in E, $u(x)-x$ is *isotropic*; this condition is obviously equivalent to the relation

$$f(x,u(x))+f(u(x),x)=0$$

or $f(x,u(x))=(f(x,u(x)))^J$. We are going to show that in such a case, the last equation is also true for every *isotropic* vector x in E, unless $K=\mathfrak{F}_4$ and $J\neq 1$ (in which case the field K_0 invariant by J is the field $\mathfrak{F}_2$ of two elements).

Let a be an arbitrary isotropic vector in E; there exists a second isotropic vector b such that $f(a,b)=1$. Let $x=a\zeta+b$; if we write that x is isotropic, we obtain the relation $\zeta+\zeta^J=0$, in other words, ζ must be a *symmetric* element. By assumption,

$$f(a\zeta+b,u(a)\zeta+u(b))=\zeta^Jf(a,u(a))\zeta+\zeta^Jf(a,u(b))+f(b,u(a))\zeta+f(b,u(b))$$

is a *symmetric* element when ζ is *not symmetric*; if we write $f(a,u(a))=\alpha$, $f(b,u(b))=\beta$, $f(a,u(b))=\lambda$, $f(b,u(a))=\mu$ we can also say that we have the relation

$$\zeta^J(\alpha+\alpha^J)\zeta+\zeta^J(\lambda+\mu^J)+(\mu+\lambda^J)\zeta+(\beta+\beta^J)=0 \tag{6}$$

for every *nonsymmetric* ζ. Now, if ζ is not symmetric, $\xi+\zeta$ is also not symmetric for every *symmetric* ξ; replacing ζ by $\xi+\zeta$ in (6) and subtracting from (6), we get

$$\xi(\alpha+\alpha^J)\xi+\zeta^J(\alpha+\alpha^J)\xi+\xi(\alpha+\alpha^J)\zeta+\xi(\lambda+\mu^J)+(\mu+\lambda^J)\xi=0 \tag{7}$$

for every nonsymmetric ζ and every symmetric ξ. Replacing now ξ by $\xi+\eta$, where ξ and η are both symmetric, and subtracting, from the resulting equation, equation (7) and the analogue with ξ replaced by η, we obtain

$$\xi(\alpha+\alpha^J)\eta+\eta(\alpha+\alpha^J)\xi=0 \tag{8}$$

for every pair of symmetric elements ξ, η. In particular, for $\eta = 1$, this shows that $\alpha + \alpha^J$ commutes with every symmetric element of K. But as K has characteristic 2, symmetric elements generate K over its center Z, ([4], p. 367, Lemma 1) hence $\alpha + \alpha^J$ must be in Z. If $\alpha + \alpha^J \neq 0$, (8) shows that any two symmetric elements are permutable, hence K is *commutative*. But then (7) is an equation of degree 2 in ξ, where the coefficient of ξ^2 is not 0, hence it has at most two distinct roots in K, and if the field K_0 of symmetric elements of K has more than 2 elements, we thus reach a contradiction, which proves that $\alpha + \alpha^J = 0$; in other words $f(a, u(a))$ is symmetric for every isotropic vector a.

This being proved, we have now the identity

$$f(x + y\xi, u(x) + u(y)\xi) = f(u(x) + u(y)\xi, x + y\xi)$$

for every pair of vectors x, y and every scalar $\xi \in K$. Writing $\lambda = f(x, u(y))$, $\mu = f(y, u(x))$, this shows that $\lambda\xi + \xi^J\mu$ is symmetric for all $\xi \in K$, in other words $(\lambda + \mu^J)\xi = \xi^J(\mu + \lambda^J)$ for all $\xi \in K$; this can only be the case if $\lambda + \mu^J = 0$, otherwise it would give, with $\rho = \lambda + \mu^J$, first $\rho^J = \rho$ (with $\xi = 1$) and then $\xi^J = \rho\xi\rho^{-1}$ for all $\xi \in K$. The involution J would thus be an automorphism of K, which is only possible if K is commutative, and $J = 1$, a case we have excluded.

From these results, we conclude therefore that the exceptional case occurs only when

$$f(x, u(y)) = (f(y, u(x)))^J \tag{9}$$

or equivalently

$$f(x, u(y)) = f(u(x), y) \tag{10}$$

for every pair of vectors x, y in E. Replacing x by $u(x)$ in (10) and taking into account the relation $f(u(x), u(y)) = f(x, y)$, we get

$$f(u^2(x) - x, y) = 0 \tag{11}$$

for all x and y, hence $u^2(x) = x$ for all $x \in E$, since f is nondegenerate. In other words, u is an *involution* in $U_n(K, f)$; as K has characteristic 2, this means that $u(x) = x + t(x)$, where t is a linear mapping of E onto a totally isotropic subspace V, $V^* \supset V$ being the kernel of t. In order to prove Theorem 2, we argue as follows: if we can choose a quasi-symmetry v such that uv is not an involution any more, then uv is a product of quasi-symmetries, and so is therefore u. Let a be a vector in V; if $u(v(a)) \neq v^{-1}(u^{-1}(a))$, uv will not be an involution. As $u(a) = a$, and $u(v(a)) = v(a) + b$, where $b \in V$, the relation $u(v(a)) = v^{-1}(u^{-1}(a))$ can be written $v(a) - v^{-1}(a) = b$, and

therefore we have to show that it is possible to find a quasi-symmetry v such that $v(a) - v^{-1}(a)$ is not in V. Take a non-isotropic vector c which is not orthogonal to a; the intersection of the plane $P = aK + cK$ with V is then the line aK. Let v be the quasi-symmetry of hyperplane H orthogonal to c, transforming c into $c\rho$ $(\rho \neq 0)$; let $a = c\lambda + z$, where $z \varepsilon H$ and $\lambda \neq 0$; then $v(a) = c\rho\lambda + z$, $v^{-1}(a) = c\rho^{-1}\lambda + z$, hence $v(a) - v^{-1}(a) = c(\rho - \rho^{-1})\lambda$, and this last vector will not be in V provided ρ can be chosen so that $\rho^2 \neq 1$ (i. e. $\rho \neq 1$), and $\rho^J f(c, c)\rho = f(c, c)$. But this is always possible when f is trace-valued and $J \neq 1$, ([4], p. 370), and this ends our proof when $K_0 \neq \mathfrak{F}_2$.

9. It is easy to see that the case $K_0 = \mathfrak{F}_2$, $n = 2$ is exceptional for Theorem 2. In the 2-dimensional vector space over $K = \mathfrak{F}_4$, there are then 5 lines aK through the origin; two of them, e_1K and e_2K, are orthogonal and non isotropic, the other three are isotropic. We have $f(e_1, e_1) = f(e_2, e_2) = 1$, since for a nonisotropic vector x, $f(x, x) \varepsilon K_0$ and is not 0, hence must be 1; the linear transformation exchanging e_1 and e_2 belongs therefore to $U_2(\mathfrak{F}_4)$; but it cannot be generated by quasi-symmetries, since these leave invariant each line e_1K, e_2K.

We have still to prove that the theorem is true for $K = \mathfrak{F}_4$ and $n \geq 3$. It will be enough to give the proof for $n = 3$, for if $n > 3$, $u \varepsilon U_n(\mathfrak{F}_4)$, and x is a nonisotropic vector in E, the two vectors x and $u(x)$ belong to a nonisotropic 3-dimensional subspace V: this is immediate if $u(x) = x$, or if the plane P through x and $u(x)$ is non isotropic. On the other hand, if P is isotropic, it is not totally isotropic, since it contains the nonisotropic vector x; there is therefore only one isotropic line aK in P, orthogonal to P; if b is a nonisotropic vector in E, not orthogonal to a, the 3-dimensional subspace $V = P + bK$ is nonisotropic, for if we write that $y = a\alpha + b\beta + x\gamma$ is orthogonal to a, b and x, we find $f(a, b)\beta = 0$, $f(b, a)\alpha + f(b, b)\beta + f(b, x) = 0$ and $f(x, b)\beta + f(x, x)\gamma = 0$; as $f(a, b) \neq 0$, $\beta = 0$, and as $f(x, x) \neq 0$, $\gamma = 0$; then the second equation reduces to $f(b, a)\alpha = 0$, which shows that $y = 0$ and proves our contention. The theorem being supposed to be true for $n = 3$, there exists *in* V a product of quasi-symmetries transforming x into $u(x)$ (since by Witt's theorem there is a unitary transformation in V transforming x into $u(x)$); these quasi-symmetries in V can be extended to quasi-symmetries *in* E by taking them to be the identity in the orthogonal subspace V^*. Then, the inductive process used in section 8 can again be applied for $n > 3$, and the proof is thus reduced to the case $n = 3$.

10. To prove the theorem in that case, we will first show that if a and b are any two nonisotropic vectors in E, there is a product of quasi-symmetries

transforming a into b. This result is obvious if $b = a\alpha$ with $\alpha \neq 0$ (and then necessarily $\alpha\alpha^J = 1$, an equation which is verified by every element $\alpha \neq 0$ in $\mathfrak{F}_4$); for then we merely consider the quasi-symmetry s of hyperplane orthogonal to a, and such that $s(a) = a\alpha = b$. Suppose next a and b are not collinear and consider an orthogonal basis e_1, e_2, e_3 of E such that $e_1 = a$; the only nonisotropic lines orthogonal to e_1K are e_2K and e_3K; the other 9 nonisotropic lines in E are in the 3 isotropic planes through e_1K and the 3 isotropic lines in the plane $e_2K + e_3K$ (3 such lines distinct from e_1K in each of these planes). Suppose first that b is not orthogonal to a; then bK is in one of the isotropic planes through e_1K, say P. Let cK be one of the nonisotropic lines in P distinct from aK and bK; as there are 3 elements $\alpha \in \mathfrak{F}_4$ such that $\alpha\alpha^J = 1$, there are 3 distinct quasi-symmetries of hyperplane H orthogonal to cK, transforming c into the 3 vectors $c\alpha$, and leaving elementwise invariant the (unique) isotropic line D in P, orthogonal to c. As such a quasi-symmetry cannot leave invariant any line in P other than cK and D, the 3 preceding quasi-symmetries transform aK into the 3 lines in P distinct from cK and D; in particular there is one such quasi-symmetry transforming aK into bK, and we have then proved our assertion.

11. There remains the case in which a and b are orthogonal; we can suppose for instance that $a = e_1$, $b = e_2$. We are going to prove slightly more, namely that there is a product of quasi-symmetries which *exchanges* e_1K *and* e_2K. Let D be an isotropic line in the plane $e_1K + e_2K$, and let cK be a nonisotropic line distinct from e_3K in the isotropic plane $P = D + e_3K$. As cK is not orthogonal to e_1K, the plane $Q = e_1K + cK$ is isotropic; there is therefore a quasi-symmetry s of hyperplane orthogonal to c, which will transform the line e_1K into another nonisotropic line xK in the plane Q. This line xK is neither in the plane $e_1K + e_2K$ (for the intersection of Q with that plane is e_1K), nor in the plane $e_3K + e_2K$ (for the intersection of that plane with Q is isotropic). Therefore the plane $R = e_2K + xK$ is isotropic. The intersection $R \cap P = dK$ is then nonisotropic, since the only isotropic line in P is D in the plane $e_1K + e_2K$; and the argument in section 10 shows that there is a quasi-symmetry s' of hyperplane orthogonal to dK, which will transform xK into e_2K. This shows that $s's$ transforms e_1K into e_2K; but in addition, both s and s' leave invariant the isotropic line D, hence $s's$ transforms the plane $e_1K + e_2K = e_1K + D = D + e_2K$ into itself; but it then transforms e_2K, which is orthogonal to e_1K in the plane $e_1K + e_2K$, into the line orthogonal to e_2K in that plane, that is e_1K.

We can now end the proof in the following way: if x is a nonisotropic vector in E, there is a product v of quasi-symmetries which transforms $u(x)$ into x; $vu = u_1$ then leaves invariant x, hence also the plane H orthogonal to x in E; let aK and bK be the nonisotropic lines (orthogonal to each other) in H. If u_1 leaves invariant both these lines, it is a product of quasi-symmetries; if it exchanges aK and bK, then there is a product w of quasi-symmetries in E which exchanges aK and bK, and therefore leaves invariant xK; the product wu_1 then leaves invariant xK, aK and bK, and is therefore a product of quasi-symmetries. Theorem 2 is thus completely proved.

12. The assumptions in this section are those of section 8, with the additional condition that $J \neq 1$; the orthogonal and symplectic groups are thus excluded. Let ϕ be the natural homomorphism of K^* onto its factor group K^*/C, where C is the commutator subgroup of K^*.

THEOREM 3. *For every unitary transformation $u \in U_n(K, f)$, the determinant* [1] *of u has the form $\phi(\gamma^J\gamma^{-1})$.*

Theorem 2 reduces the proof of Theorem 3 to the case in which u is a quasi-symmetry (the exceptional case of Theorem 2 is of course known to satisfy also Theorem 3, which is well known for commutative fields K). Let us therefore suppose that u is a quasi-symmetry of hyperplane H, and let e be a vector orthogonal to H, and $\rho = f(e, e) \neq 0$; then $u(e) = e\alpha$ with $\alpha^J\rho\alpha = \rho$, and the determinant of u is obviously $\phi(\alpha)$ (take as a basis of E the vector e and a basis of H). We are thus reduced to prove that the relation $\alpha^J\rho\alpha = \rho$ implies $\alpha = \gamma^J\gamma^{-1}$ for some $\gamma \neq 0$. This is immediate if $\alpha = -1$, for if λ is an element in K such that $\lambda^J \neq \lambda$, then for $\gamma = \lambda - \lambda^J$, $\gamma^J = -\gamma$, and as $\gamma \neq 0$, $-1 = \gamma^J\gamma^{-1}$. If $\alpha \neq -1$, consider the element $\gamma^{-1} = \rho(1 + \alpha)$; we have $\gamma^{-J} = (1 + \alpha^J)\rho$, hence $\gamma^{-J}\alpha = \rho\alpha + \alpha^J\rho\alpha = \rho\alpha + \rho = \gamma^{-1}$, whence $\alpha = \gamma^J\gamma^{-1}$, and this ends our proof.

13. We add two disconnected remarks on results proved in [4] and [2] respectively. The first concerns Theorem 4 of [4], p. 380 which is proved in that paper for a sfield K of characteristic $\neq 2$. I want to show that the theorem is still valid when that restriction on the characteristic is dropped. To prove this, let us first remark that we can suppose that there is a vector $a \in E$ such that $f(a, a) = 1$ (excluding the symplectic groups, whose structure is well known). Indeed, let a be a nonisotropic vector, and let $f(a, a) = \alpha \neq 0$; α is a symmetric element for the involution J, the form $f_1(x, y) = \alpha^{-1}f(x, y)$ is hermitian for the involution $\xi \to \xi^T = \alpha^{-1}\xi^J\alpha$; moreover, f_1 is trace-valued

for that involution, for we have $\alpha^{-1}(\xi + \xi^J) = (\alpha^{-1}\xi) + (\alpha^{-1}\xi)^T$; finally, $f_1(a, a) = 1$, and it is clear that the unitary groups $U_n(K, f)$ and $U_n(K, f_1)$ are identical. With this assumption, we have therefore $1 = \rho + \rho^J$; Theorem 4 of [4] will then be proved, if we can prove Lemma 6, [4], p. 380, and this in turn amounts to finding 4 elements α, β, α_1, β_1 in K such that $\alpha_1\alpha_1{}^J = \alpha\alpha^J$, $\beta_1\beta_1{}^J = \beta\beta^J$, and $\alpha_1\beta_1{}^J - \alpha\beta^J = \gamma$, where γ is an arbitrarily given element of K. In order to show that this is possible, take $\alpha_1 = \alpha$, and $\beta_1 = 1$; if we can prove that β can be chosen such that $\beta\beta^J = 1$ and $\beta \neq 1$, the equation $\alpha_1\beta_1{}^J - \alpha\beta^J = \gamma$ reduces to $\alpha(1 - \beta^J) = \gamma$ and always has a solution. But if we take $\beta = \rho^J\rho^{-1}$, then $\beta\rho\beta^J = \rho$, $\beta\rho^J\beta^J = \rho^J$, hence $\beta\beta^J = \beta(\rho + \rho^J)\beta^J = \rho + \rho^J = 1$, and $\rho^J \neq \rho$, since otherwise $\rho + \rho^J$ would be 0.

14. The other remark is relative to the proof, given in [2], pp. 72-73, that every noncommutative sfield K for which every element has degree ≤ 2 over the center Z of K, is a reflexive sfield. The first remark in the proof is that every finite set $(\xi_i)_{1 \leq i \leq n}$ of elements of K generates over Z a sfield $Z[\xi_1, \cdots, \xi_n]$ of rank $\leq 2^n$; but this does not (contrary to what is asserted in [2], p. 73) reduce the problem to the case in which $[K:Z]$ is finite, for the center of $K[\xi_1, \cdots, \xi_n]$ might be distinct from Z. To fill that gap in the proof, let us consider two nonpermutable elements ξ, η of K; as the rank $[Z[\xi, \eta]:Z]$ is ≤ 4 and is a multiple of 2 which cannot be equal to 2 (otherwise $Z[\xi, \eta]$ would be commutative) it is equal to 4. Let now ζ be any other element of K, and let T be the center of $Z[\xi, \eta, \zeta]$; we want to prove that $T = Z$, for then the argument of ([2], p. 73) shows that $Z[\xi, \eta, \zeta]$ has rank 4 over Z, hence $\zeta \in Z[\xi, \eta]$, and therefore $K = Z[\xi, \eta]$. Suppose $T \neq Z$, hence as $[Z[\xi, \eta, \zeta]:Z] \leq 8$, we have $[Z[\xi, \eta, \zeta]:T] < 8$, and as the rank of $Z[\xi, \eta, \zeta]$ over its center T must be a square and cannot be 1, $[Z[\xi, \eta, \zeta]:T] = 4$, and as $T \neq Z$, $[T:Z] = 2$. There is then a maximal subfield S of $Z[\xi, \eta, \zeta]$ such that $[S:T] = 2$ and that S is separable over T. We can write therefore $S = T(\alpha)$, where α is separable and of degree 2 over T; but α is also of degree 2 over $Z \subset T$, hence is separable over Z. On the other hand, we have $T = Z(\beta)$, where β has degree 2 over Z, but might be inseparable when Z has characteristic 2. Nevertheless, in such circumstances, it is known ([5], p. 132) that for the field $S = Z(\alpha, \beta)$, the theorem of the primitive element still holds, in other words $S = Z(\gamma)$, where γ has degree 4 over Z; but this contradicts the assumption that the degree of γ over Z is at most 2, and the assumption $T \neq Z$ is therefore untenable, which ends the proof.

UNIVERSITY OF MICHIGAN.

BIBLIOGRAPHY.

[1] J. Dieudonné, "Les déterminants sur un corps non commutatif," *Bulletin de la Société Mathématique de France*, vol. 71 (1943), pp. 27-45.

[2] ———, "Sur les groupes classiques," *Actualités Scientifiques et industrielles*, no. 1040, Paris (Hermann), 1948.

[3] ———, "On the automorphisms of the classical groups," *Memoirs of the American Mathematical Society*, no. 2 (1951).

[4] ———, "On the structure of unitary groups," *Transactions of the American Mathematical Society*, vol. 72 (1952), pp. 367-385.

[5] B. L. van der Waerden, *Moderne Algebra*, t. I (2nd ed.); Berlin (Springer), 1937.

ON THE LOCAL RÔLE OF THE THEORY OF THE LOGARITHMIC POTENTIAL IN DIFFERENTIAL GEOMETRY.*

By Aurel Wintner.

Definitions. In a (u, v)-plane, let D be an open domain (which, for the purposes at hand, can be assumed to be simply connected and "sufficiently small") and let $g_{11}, g_{12} = g_{21}, g_{22}$ be three functions which are of class C^n and satisfy the conditions $g_{ii} > 0$, $\det g_{ik} > 0$ on D. Then

$$g_{\alpha\beta}(u^1, u^2)\, du^\alpha du^\beta, \quad \text{where} \quad u^1 = u, u^1 = v, \tag{1}$$

will be called a C^n-metric (on D).

Let

$$g'_{\alpha\beta}(u'^1, u'^2)\, du'^\alpha du'^\beta, \quad \text{where} \quad u'^1 = u', u'^2 = v', \tag{2}$$

be a C^m-metric on a (u', v')-domain D'. Then the C^n-metric (1) is called isometric to (2) if, corresponding to every point (u_0, v_0) of D, there exist a circle

$$D_0\colon\ (u - u_0)^2 + (v - v_0)^2 < \epsilon^2 \tag{3}$$

and a C^1-mapping

$$u' = u'(u, v), \qquad v' = v'(u, v) \tag{4}$$

of (3) on a (*schlicht*) domain D'_0 in D' in such a way that (1) becomes identical with (2) on D_0 (or D'_0) by virtue of (4) (or

$$u = u(u', v'), \qquad v = v(u', v'), \tag{5}$$

the inverse of (4)). It is understood that (4) is called a (local) C^r-mapping, where $r \geqq 1$, if both functions (4) are of class C^r and have a non-vanishing Jacobian (which implies that, if (3) is small enough, D'_0 is *schlicht* and (5) is a C^r-mapping of D'_0 onto D_0).

If (1) is a C^n-metric, where $n \geqq 2$, then it has a curvature $K = K(u, v)$, defined, if

$$g = (\det g_{ik})^{\frac{1}{2}} \qquad (g > 0), \tag{6}$$

---

* Received April 25, 1953.

by

$$(7) \qquad -2gK = \{(g_{22u} - g_{12v})/g\}_u + \{(g_{11v} - g_{12u})/g\}_v$$
$$+ (2g^3)^{-1}\det(\Gamma, \Gamma_u, \Gamma_v), \quad \text{where} \quad \Gamma = (g_{11}, g_{12}, g_{22})$$

(the subscripts u and v denote partial differentiations). Thus $K(u, v)$ is a function of class C^{n-2} if (1) is a C^n-metric, where $n \geqq 2$.

In particular, (1) must be a C^3-metric before it is *sure* to have a $K(u, v)$ of class C^1. On the other hand, $K(u, v)$ *can* have this property also if (1) is just a C^2-metric. Actually, it turns out that this type of a metric, a C^2-metric (1) for which the curvature $K(u, v)$ is a function of class C^1 (in terms of the parameters (u, v) in which the coefficients $g_{ik}(u, v)$ of (1) are given as functions of class C^2), is quite an important class. In fact, it will be shown that for *this* class of metrics there prevail simple facts (two of which, Theorem 1 and Theorem 3, become false if they are referred to either of the, more general or less general, classes of C^n-metrics, where $n = 2, 3$).

Conformal normal forms. Let a metric (1) be called *conformal* (that is to mean, "conformal with reference to the euclidean (u, v)-plane," i. e., "isothermic") if

$$(8) \qquad g_{11} = g_{22} \quad \text{and} \quad g_{12} = 0$$

hold identically in (u, v); so that (1) reduces to

$$(9) \qquad g(u, v)(du^2 + dv^2), \qquad (g > 0),$$

if g denotes the common value of g_{11} and g_{22} (in view of (8), this agrees with the more general notation (6)). According to Lichtenstein [4], every C^1-metric (1) is isometric to a conformal metric

$$(10) \qquad g'(u', v')(du'^2 + dv'^2) \qquad (g' > 0).$$

In fact, he proved that this is true also if the coefficients $g_{ik}(u, v)$ of (1), instead of being functions of class C^1, are only of class $C^{\epsilon-0}$ (for some $\epsilon > 0$), where $C^{\lambda-0}$ denotes (for $0 < \lambda \leqq 1$) the class of those functions which belong for every $\mu < \lambda$ to the class C^μ of those functions which satisfy a locally uniform Hölder condition of index μ. On the other hand, it was shown in [2] that not every metric (1) with continuous coefficients $g_{ik}(u, v)$ is isometric to a (continuous) conformal metric.

While Lichtenstein's method [4] also proves that every C^2-metric is isometric to a conformal C^1-metric, it was shown in [2] that not every

C^2-metric is isometric to a conformal C^2-metric. In order that a metric be isometric to a conformal C^2-metric, it is sufficient to assume that the former is a C^3-metric but it is not true that every C^3-metric is isometric to a conformal C^3-metric; cf. [4] and [2], respectively. It will be shown that this disorder can be disposed of by considering the class of C^2-metrics possessing a curvature of class C^1, a class intermediary between the class of the C^2-metrics and that of the C^3-metrics.

THEOREM 1. *Every C^2-metric* (1) *possessing a curvature $K(u,v)$ of class C^1 is isometric to a conformal C^2-metric* (10) *possessing a curvature $K'(u',v')$ of class C^1 (and every C^1-mapping* (4) *establishing the isometry must be a C^3-mapping).*

Accordingly, every C^2-metric possessing a curvature $K(u,v)$ of class C^1 (but, as mentioned before, not an arbitrary C^2-metric, the curvature of which is just continuous in general) must be isometric to a conformal C^2-metric. Actually, the latter statement is not weaker than Theorem 1, since, the curvature

$$K'(u',v') = K(u,v) = K(u(u',v'), v(u',v')) \tag{11}$$

being a scalar, its C^1-character is preserved by every C^1-mapping (4). Correspondingly, the parenthetical assertion of Theorem 1 is not deeper than (and is of course contained in) in the *second* of the following statements (I), (II), $\cdots$:

(I) If (1) and (10) are isometric C^1-metrics, then every C^1-mapping (4) establishing their isometry must be a C^2-mapping.

(II) The assertion of (I) remains true if C^1 and C^2 are replaced by C^2 and C^3, respectively. Etc.

(I), (II), $\cdots$ can be proved in various ways. The significance of these and of more general conclusions was pointed out in [8], pp. 199-203. The most general result in the binary and definite case was proved by Hartman [1].

The simplest proof of (I), (II), $\cdots$ seems to be a reduction to the particular case in which (10) is the euclidean metric ($g' = 1$). Let, (I*), (II*), $\cdots$ denote the particular statements which thus result from (I), (II), $\cdots$, respectively (that is, the statements in which

$$du'^2 + dv'^2 \tag{12}$$

replaces (10) in (I), (II), $\cdots$). For a proof of (I*), which is valid for any dimension number and for indefinite metrics also, cf. [7], pp. 571-573.

2

The proof of (II*) is the same as that of (I*) (but in one respect simpler; cf. *loc. cit.*). But the particular cases (I*), (II*) of (I), (II) imply the latter. This can be seen as follows: Under the assumptions and in the notations of (I), define a (positive) function f of (u,v) by placing

$$(13) \qquad f(u,v)=1/g'(u'(u,v),v'(u,v)).$$

Then $f(u,v)$ is of class C^1, since $g'(u',v')$ and both functions (4) are. According to (13), the isometry of (10) and (1) means the isometry of the euclidean metric (12) and of the C^1-metric

$$(14) \qquad f(u^1,u^2)g_{\alpha\beta}(u^1,u^2)du^\alpha du^\beta, \quad \text{where} \quad u^1=u,\ u^2=v.$$

Hence, the assertion of (I) follows by applying (I*) to (14) and (12). Similarly, (II) follows from (II*) if use is made of (I).

There remains to be proved the main assertion of Theorem 1. The proof will be such as to supply the following results as a by-product:

THEOREM 2. *If the curvature $K(u,v)$ of a C^2-metric* (1) *is a function of class C^j, and if $j\leqq 3$, then* (1) *is isometric to a C^{j+1}-metric.*

Incidentally, the latter metric will result as a *conformal* C^{j+1}-metric.

One might expect that some information on the C^2-metric will also result if the C^j-character of $K(u,v)$ is required for an index $j\geqq 4$. The place assigned to C^2-metrics (1) by the scale (I), (II), $\cdots$ seems, however, to indicate that the process of refining is arrested at $j=3$.

That scale prevents, in particular, the extension of Theorem 2 to a scale the next item of which would assert that, if $j=4$, the C^2-metric must be isometric to a C^5-metric (but a counterexample is missing).

Proof of Theorem 1. Let (1) be a C^2-metric on a (u,v)-domain D. Then the function K of (u,v) is just continuous, while the functions (6) and Γ, occurring in (7), are of class C^2 and C^1, respectively. Hence, if $B=B(J)$ denotes the interior of a positively oriented, rectifiable Jordan curve J contained in D, then an integration of (7) over B, when followed by an application of Green's formula, leads to the following identity in $J=J(B)$:

$$(15) \qquad \int_J (2g)^{-1}\{g_{11v}-g_{12u})du+(g_{12v}-g_{22u})\}dv$$

$$=\int\!\!\int_B g\{K+(4g^4)^{-1}\det(\Gamma,\Gamma_u,\Gamma_v)\}dudv, \quad \text{where} \quad \Gamma=(g_{11},g_{12},g_{22}).$$

Actually, the existence of a continuous function $K = K(u, v)$ satisfying this integral identity can be thought of as defining a C^1-metric (1) which possesses a continuous curvature (note that (7) does not apply to a metric which is just of class C^1). This idea is due to Weyl [5], pp. 42-44.

Suppose that the C^2-metric (1) occurring in (15) is transformed by a C^1-mapping (4) into a metric (2) which, for some reason, is known to be a C^1-metric (rather than just a continuous metric) on the (u', v')-image D' of D. Then, according to [1], p. 222, the C^1-mapping (4) must be a C^2-mapping. Denote by (15′) the relation which results if u, v, g_{ik}, g, Γ, K (and J, B) are replaced by u', v', g'_{ik}, g', Γ', K' (and the (u', v')-images J', B' of J, B), respectively. Then it can be verified from the definitions $g = (\det g_{ik})^{\frac{1}{2}}$, $g' = (\det g'_{ik})^{\frac{1}{2}}$ and from the tensor character of (g_{ik}), (g'_{ik}) that, if (11) is considered as the definition of $K'(u', v')$, the relation (15′) holds as an identity in J' or B', where $B' = B'(J')$ denotes the interior of any positively oriented, rectifiable Jordan curve contained in D'. It follows (if (1) and (2) are interchanged) that the identity (15) in J holds not only for C^2-metrics but also for every C^1-metric which (for some reason) is isometric to a C^2-metric.

As explained after Theorem 1, the latter is equivalent to the statement that every C^2-metric (2) possessing a curvature $K'(u', v')$ of class C^1 is isometric to a conformal C^2-metric (2), that is, to a metric (9) in which $g(u, v)$ is a function of class C^2. In order to prove the latter formulation of Theorem 1, assume first only that the given metric (2) is a C^2-metric. According to a theorem of Lichtenstein, mentioned above, any such metric (2) is isometric to a conformal C^1-metric; that is, to a metric (9) in which $g(u, v)$ is a function of class C^1. Since the C^1-metric (9) is isometric to a C^2-metric, (15) is applicable to (9) and reduces, in view of (8), to

$$\int_J (2g)^{-1}(g_v du - g_u dv) = \iint_B gK\,du\,dv. \tag{16}$$

Hence, in order to prove Theorem 1, it remains to be shown that if not only $g(u, v)$ but also $K = K(u, v)$ (cf. (11) and (4)) is of class C^1, an assumption of Theorem 1 which was not used thus far, then (16) implies that $g(u, v)$ is of class C^2.

To this end, note that if

$$f(u, v) = -2K(u, v)g(u, v) \tag{17}$$

and

$$\gamma(u, v) = \log g(u, v), \qquad (g > 0), \tag{18}$$

then (16) can be written in the form

$$(19) \qquad \int_J (\gamma_u dv - \gamma_v du) = \iint_B f(u,v)\,du dv$$

and that (19) is just the integrated form of Poisson's equation

$$(20) \qquad \gamma_{uu} + \gamma_{vv} = f(u,v).$$

Correspondingly, the proof of Theorem 1 must be made to depend on an appeal to the theory of the logarithmic potential

$$(21) \qquad \phi(u,v) = \tfrac{1}{4\pi}\iint_A f(x,y)\log\{(u-x)^2 + (v-x)^2\}\,dxdy.$$

It is known that if $f(u,v)$ is an arbitrary continuous function (not satisfying anything like a Hölder condition), then (i) the differential equation (20) need not possess any (continuous) solution $\gamma(u,v)$ but (ii) the integrated form (19) of (20) is satisfied by the sum, $\gamma(u,v)$, of any regular harmonic function $h(u,v)$ and the logarithmic potential (21), which is a function $\phi(u,v)$ of class C^1, finally (iii) every function γ of class C^1 satisfying (19) is of the form $h+\phi$; cf. [9], [3], [6]. It is understood that the continuous function $f(u,v)$ is supposed to be given on a (u,v)-domain D and that, if A is any bounded domain the closure of which is contained in D, then, in the above assertions, (21) is considered only for the domain of points (u,v) which constitute A.

Since $g(u,v)$ and (by assumption) $K(u,v)$ are functions of class C^1, the same is true of the function $f(u,v)$ defined by (17). Hence the corresponding logarithmic potential (21) is a function of class C^2 on A (the conclusion that (21) is of class C^2 on A, whenever $f(u,v)$ is of class C^1, is due to Gauss; the finer aspects of Hölder's theory are not needed). Since every function γ of class C^1 satisfying (19) is of the form $h+\phi$, it follows that every such γ is of class C^2. Finally, since (18) is a (particular) function of class C^1 satisfying (19), it follows that the γ occurring in (18), and therefore $g = e^\gamma$ as well, is of class C^2.

This proves Theorem 1.

Proof of Theorem 2. If $j=1$, then Theorem 2 and the remark made after it are contained in Theorem 1. The remaining two cases, $j=2$ and $j=3$, must be dealt with in succession.

Suppose first that the assumptions of the case $j=2$ of Theorem 2 are satisfied. Then, in view of Theorem 1 (including its parenthetical assertion; cf. (11)), it can be assumed that the given metric is of the type (9). Thus both g and K are now of class C^2, and the assertion is that g must therefore be of class C^3.

Corresponding to the result of Gauss, used above, the logarithmic potential (21) of every density f of class C^2 is of class C^3. Since the product (17) is of class C^2, it follows, by the argument applied at the end of the proof of Theorem 1, that g is of class C^3.

This proves Theorem 2, as well as the remark made after Theorem 2, for the case $j=2$. The corresponding assertion for the case $j=3$ follows in the same way as in the case $j=2$ (if use is made of results proved for the latter case).

General indices. All of this can be summarized as the case $n=2$ of the following theorem (*):

(*) *If* (1) *is a C^n-metric, where $n>1$, and if* (1) *has a curvature $K(u,v)$ of class C^{n-1+m}, where $0 \leqq m \leqq 2$, then* (1) *is isometric to a conformal C^{n+m}-metric* (10) (*and K is a function of class C^{n-1+m} in terms of the parameters u', v' of* (10) *also*).

Clearly, the above proofs, given for $n=2$, hold for $n=3,4,\cdots$ also. The necessity of restricting m to its three lowest values remains undecided (for every n); cf. the remarks made after Theorem 2.

With regard to the case $n=1$, excluded in the wording of (*), it is clear from the proof of Theorem 2 that if (*) is true for $(n,m)=(1,0)$, then it is true for $(n,m)=(1,1)$ and $(n,m)=(1,2)$ also (here a C^1-metric (1) possessing a $K(u,v)$ of class C^0 is defined in the sense of Weyl, as explained after (15)). But the truth of (*) for the case $(n,m)=(1,0)$ remains undecided.[1]

[1] Weyl has claimed ([5], p. 49) that the truth of (*) for $(n,m)=(1,0)$ is known from the theory of functions (in his case of a closed convex surface, where $K>0$). Actually, all that follows by function-theoretical methods is that if the theorem is true in the small, then (subject to trivial topological restrictions) it is true in the large also. The sharpest known results in the small seem to be those of Lichtenstein, referred to above.

An apparent exception seems to be contained between the lines of a proof of a (non-local) theorem of Weyl [5], p. 68, in which the metric (1), instead of satisfying the assumptions of Theorem 1, is merely required to be a C^1-metric possessing a curvature $K(u,v)$ which, along with the first derivatives of the coefficients of (1), is sub-

Surfaces. By a surface of class C^n, where $n \geqq 1$, will be meant a set S of points $X = (x, y, z)$ having the property that, in a neighborhood of every point of it, S can be parametrized in the form

$$X = X(u, v), \qquad (X = (x, y, z)), \tag{22}$$

where (u, v) varies on some, sufficiently small, domain D, the vector function (22) is of class C^n and the vector product, say $V = V(u, v)$, of the partial derivatives X_u, X_v does not vanish; so that $N = V/|V|$ defines a unit vector which is a function $N(u, v)$ of class C^{n-1}. Then (22) will be called a C^n-parametrization of the surface S of class C^n. Note that, if $n > m \geqq 1$, every surface S of class C^n possesses C^m-parametrizations which are not C^{m+1}-parametrizations.

If (22) is of class C^3, then, since $V \neq 0$, the first fundamental form

$$|dX(u^1, u^2)|^2 = g_{\alpha\beta}(u^1, u^2)du^\alpha du^\beta \qquad (u^1 = u, u^2 = v) \tag{23}$$

on S is a metric (1) of class C^2. But more than this is true:

THEOREM 3. *In order that a C^2-metric* (1) *be realizable as the metric* (23) *on some surface* (22) *of class C^3, it is necessary for* (1) *to have a curvature $K(u, v)$ of class C^1.*

In fact, if $n \geqq 2$, then, since the unit vector $N(u, v)$ is of class C^{n-1}, the second fundamental form on S, being defined as the scalar product of $-dN(u, v)$ and $dX(u, v)$, exists and has coefficients, say $h_{ik} = h_{ki}$, which are functions of class C^{n-2}. Since the coefficients of (23) are functions of class C^{n-1} and since, due to the embedding (22)-(23) of the metric (1) on S, the ratio of the determinants $\det h_{ik}(u, v)$, $\det g_{ik}(u, v) > 0$ is identical with the curvature $K(u, v)$ of the metric (Gauss), it follows that $K(u, v)$ is of class C^{n-1} if $n > 2$, and exists and is continuous if $n = 2$ (Weyl; cf. (15)). Theorem 3 is the case $n = 3$ of this conclusion, which applies the assertion of the *theorema egregium* in the direction just opposite to that emphasized by Gauss.

Let a C^n-parametrization (22), where $n \geqq 1$, be called an *isothermic parametrization of a surface S* if (8) holds for the first fundamental form (23). Then it follows from Theorem 3 that Theorem 1 (along with the

jected to a Hölder condition (*loc. cit.*, K is positive on a closed orientable (u, v)-manifold). It turns out, however, that any such metric must be isometric to a C^2-metric and, what is more, to a Hölderian C^2-metric. This is the first assertion of Theorem 5 below; cf. also Theorem 6.

parenthetical assertion of Theorem 1) contains the following fact as a particular case:

THEOREM 4. *Every surface S of class C^3 possesses isothermic C^3-parametrizations.*

The above proofs make it clear that Theorem 4 remains true if both of its classes C^3 are replaced by C^n, provided that $n > 2$ (and it seems to be likely that $n = 2$ can here be included), whereas Lichtenstein's theory [4] supplies only the existence of an isothermic parametrization of class C^{n-1} if the surface S is of class C^n, where $n = 2, 3, \cdots$. This shows the "natural" character of the assumptions and assertions of Theorem 1 (and of its analogues for other n-values).

Hölder restrictions. Let a function $f(u, v)$, defined on a (u, v)-domain D, be called of class HC^n, where $n \geqq 0$, if it is of class C^n and all of its partial derivatives $\partial^{h+j}f/\partial u^h \partial v^j$, where $h + j \leqq n$, satisfy a (locally) uniform Hölder condition of *some* positive index $\lambda (= \epsilon < 1)$. Correspondingly, a mapping (4) will be called an HC^n-mapping, where $n \geqq 1$, if it is a C^n-mapping (with non-vanishing Jacobian) and both functions u', v' occurring in (4) are of class HC^n. Similarly, a metric (1) will be called an HC^n-metric, where $n \geqq 0$, if all three functions $g_{ik}(u, v)$ are of class HC^n.

In this terminology, a variant of Theorem 1 can be formulated as follows:

THEOREM 5. *Let* (1) *be an HC^1-metric possessing a curvature $K(u, v)$ of class HC^0. Then* (1) *is isometric to an HC^2-metric. The latter can be chosen to be a conformal metric. Finally, the mapping* (4)-(5), *establishing the isometry of the given metric* (1) *and the conformal HC^2-metric* (10), *is realized by an HC^2-mapping.*

The same is true if HC^1, HC^0, HC^2 are replaced by HC^n, HC^{n-1}, HC^{n+1}, respectively, where $n \geqq 1$.

The implications of this theorem are revealed by the following fact:

THEOREM 6. *Every conformal C^n-metric possessing a curvature of class HC^{n-1} is an HC^{n+1}-metric, where $n \geqq 1$.*

It can readily be concluded from the results of Lichtenstein, quoted before Theorem 1, that the case $n = k \geqq 1$ of Theorem 5 follows if Theorem 6 is proved for the case $n = k + 1 \geqq 2$. But Theorem 6 holds for every $n \geqq 1$.

In order to see this, let $n=1$ (if $n>1$, the proof is the same, though more straightforward). Then the assumption of Theorem 6 is that a function $g(u,v)>0$ of class C^1 and a function $K(u,v)$ of class HC^0 satisfy (16) as an identity in $B=B(J)$. In view of (17) and (18), this means that $\gamma(u,v)$ is a function of class C^1 satisfying (19), where $f(u,v)$ is a function subject to a locally uniform Hölder condition. Since (19) is the integrated form of (20), it now follows (for the reasons (i)-(iii), mentioned after (20)), that $\gamma(u,v)$ is of class C^2 and that its second derivatives satisfy a locally uniform Hölder condition. It follows therefore from (18) that $g(u,v)$ is of class HC^2, as claimed by the case $n=1$ of Theorem 6.

Appendix.*

Theorem 3 and its extensions to the classes $C^4, C^5, \cdots$ can be formulated as follows: *If a surface S has a C^n-parametrization* (22), *where $n>2$, then the C^{n-1}-metric* (23) *is isometric to a conformal C^{n-1}-metric.* It is natural to ask whether there is a corresponding theorem for the case in which the first fundamental form (23) is replaced by the second fundamental form

$$(24) \qquad -dN(u,v).\,dX(u,v)=h_{\alpha\beta}(u^1,u^2)\,du^\alpha du^\beta,$$

provided that (24) is a metric, i. e., provided that the curvature $K(u,v)$ of (22) or (23) is positive. The answer to this question turns out to be positive: *If a surface S of postive curvature K has a C^n-parametrization* (22), *where $n>3$, then the C^{n-2}-metric* (24) *is isometric to a conformal C^{n-2}-metric.* This follows from Theorem 1 and its extensions to C^k-metrics, where $k\geqq 2$, in the same way as the theorem italicized before (24) does, if the following counterpart of Theorem 3 is proved:

THEOREM 3 bis. *If a surface S of non-vanishing curvature K has a C^n-parametrization* (22), *where $n>2$, then, although the coefficients $h_{ik}(u,v)$ of* (24), *where $(u,v)=(u^1,u^2)$, are just of class C^{n-2} in general, the binary quadratic differential form* (24) (*which is definite or indefinite according as the curvature $K(u,v)$ of* (23) *is positive or negative*) *possesses a curvature $\kappa(u,v)$ of class C^{n-3}* (which means the existence of continuous $\kappa(u,v)$ in the limiting case $n=3$).

In the proof of the theorem italicized after (24), only the case $K(u,v)>0$ of Theorem 3 bis is needed (and only for $n>3$). At points (u,v) at which $K(u,v)$ vanishes, the assertion of Theorem 3 bis cannot even

* Added July 28, 1953.

be formulated, since $\kappa(u,v)$ is undefined unless either $K(u,v)>0$ or $K(u,v)<0$.

Proof of Theorem 3 bis. Suppose first that $n>3$, say $n=4$. Then the coefficients $g_{ik}(u,v)$ and $h_{ik}(u,v)$ of (22) and (23) are of class C^3 and C^2, respectively. Hence both $h_{ik}(u,v)$ and the Christoffel coefficients $\Gamma_{ik}{}^j(u,v)$ of (22) are functions of class C^2. But the equations of Codazzi-Mainardi state that the differences $h_{11v}-h_{12u}$, $h_{22u}-h_{12v}$ are certain bilinear forms in h_{ik}, $\Gamma_{ik}{}^j$. Consequently, these differences are functions of class C^2 (in (u,v)), which implies that the derivatives

$$(25) \qquad (h_{11v}-h_{12u})_v,\ (h_{22u}-h_{12v})_u$$

are functions of class C^1.

On the other hand, the representation (7) of the curvature $K(u,v)$ of the definite metric (22) shows that the principal part of $K(u,v)$, the part of $K(u,v)$ in which only the contributions of the second derivatives of the functions $g_{ik}(u,v)$ are retained, is precisely the sum of the two expressions which result if the metric $\| h_{ik} \|$ is replaced by $\| g_{ik} \|$ in (25). Correspondingly, the principal part of the curvature $\kappa(u,v)$ of (24) is the sum of the two functions (25), provided that (24) is definite ($\det h_{ik}>0$, i. e., $K>0$). Hence, under this proviso, the principal part of $\kappa(u,v)$, and therefore $\kappa(u,v)$ itself, will be of class C^1, since both functions (25) are.

This proves Theorem 3 bis for the case $n=4$ if $\det h_{ik}>0$. If $\det h_{ik}<0$, the proof remains the same, except that the definition of the curvature $\kappa(u,v)$ of an indefinite binary metric must then be used. It is also clear that, in both cases, the proof remains the same if $n=4$ is replaced by any $n>3$. Finally, the proof remains valid in the limiting case $n=3$ also, except that the equations of Codazzi-Mainardi and the definition of the curvature of a non-singular binary metric must then be replaced by their integrated forms (forms which correspond to the replacement of (7), (20) by (15), (19), respectively).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On unsmooth two-dimensional Riemannian metrics," *American Journal of Mathematics*, vol. 74 (1952), pp. 215-226.

[2] ——— and A. Wintner, "On the existence of Riemannian manifolds which cannot carry non-constant analytic or harmonic functions in the small," *ibid.*, vol. 75 (1953), pp. 260-276.

[3] L. Lichtenstein, "Ueber einige Integrabilitätsbedingungen zweigliedriger Differentialausdrücke mit einer Anwendung auf den Cauchyschen Integralsatz," *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, vol. 9 (1910), pp. 84-100.

[4] ———, "Zur Theorie der konformen Abbildung. Konforme Abbildung nicht-analytischer singularitätenfreier Flächenstücke auf ebene Gebiete," *Bulletin de l'Académie des Sciences de Cracovie*, ser. A (1916), pp. 192-217.

[5] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[6] A. Wintner, "On the Hölder restrictions in the theory of partial differential equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 731-738.

[7] ———, "On Riemann metrics of constant curvature," *ibid.*, vol. 73 (1951), pp. 569-575.

[8] ———, "On isometric surfaces," *ibid.*, vol. 74 (1952), pp. 198-214.

[9] S. Zaremba, "Contribution à la théorie d'une équation fonctionelle de la physique," *Rendiconti del Circolo Matematico di Palermo*, vol. 19 (1905), pp. 140-150.

CORRIGENDA.

Omit the factor ρ^{-1} in all three integrands on p. 273 of [2].

THE GEOMETRY OF THE LINEAR PARTIAL DIFFERENTIAL EQUATION OF THE SECOND ORDER.*

By Richard L. Ingraham.

1. Introduction. The general linear partial differential equation of the second order in one unknown and n variables has an intrinsic geometry defined by its coefficients. This was investigated first by E. Cotton [1], to whom all the basic results are due. Further work introduced little that was new. Struik and Wiener [2], who were mainly interested in a certain physical application of the Cotton theory, recognized that the geometry of the quadratic and linear differential forms involved, under the groups allowed, could be unified in the concept of one geometry—the Weyl geometry—but otherwise added nothing new to the mathematical theory. Levi-Cività [3], using the Cotton theory, confined himself to the problem of finding normal forms, eliminating one of the most interesting groups by a normalization. Moreover, he paid most attention to the case $n = 2$, an exceptional case to which the general theory does not apply.

The present paper aims first, by making consistent use of the intrinsic Weyl geometry, to cast the known theory in the form in which the powerful transformation calculus of modern differential geometry can be most directly applied to the equivalence problem (which yields a classification) and to the problems of simplifying the equation in the large by suitable transformations and of finding solutions. Second, making use of these methods, it gives several new results, of which the most important is the criterion for the equivalence of two such equations expressed in finite form in terms of complete sets of invariants of the corresponding Weyl geometries. As a corollary, the criterion that an equation be reducible to ordinary Laplacian form is immediate.

2. The intrinsic geometry. We treat only equations with vanishing undifferentiated term. Let the equation be written (cf. [4] for a concise summary of the notations used by Schouten and others)

$$(2.1)\quad g^{rs}\partial^2{}_{rs}\phi + d^r\partial_r\phi = 0 \ \ [\partial_r = \partial/\partial x^r;\ r, s, \cdots, = 1, \cdots, n;\ \det g^{rs} \neq 0].$$

* Received March 31, 1953.

The question of whether the left member might represent simply the Laplacian of ϕ in a curved space endowed with a suitable linear connection and metric is answered in the affirmative by the following theorem.

THEOREM 1. *Equations of the type (2.1) can always be written as the generalized Laplacian of ϕ equals zero in terms of covariant differentiation with respect to a unique Weyl-type linear connection. The associated intrinsic geometry of the quation is a Weyl geometry W_n $(n \neq 2)$ and is uniquely determined.*

Proof. The first part of the theorem states that (2.1) is identical with

$$\nabla_r(g^{rs}\partial_s\phi) = 0 \tag{2.2}$$

for a unique Weyl-type connection, i.e., a symmetric linear connection $\Lambda_{pq}{}^t = \Lambda_{qp}{}^t$ for which there exist a symmetric tensor $G_{rs}(\det G_{rs} \neq 0)$ and a vector F_r such that

$$\Lambda_{pq}{}^t = C_{pq}{}^t - \tfrac{1}{2}(A^t{}_pF_q + A^t{}_qF_p - G_{pq}F^t) \tag{2.3}$$

where $C_{pq}{}^t$ is the Christoffel symbol of G_{rs}, G^{rs} are the normalized cofactors, the unit tensor $A^p{}_q \equiv \delta^p{}_q$, and $F^t = G^{tr}F_r$. Expanding (2.2) using (2.3), and comparing with (2.1), one obtains the unique solutions $G_{rs} = g_{rs}$, the normalized cofactors of g^{rs}, and

$$F^r = -2/n(d^r + g^{pq}C_{pq}{}^r), \tag{2.4}$$

which proves the first part. For reasons which will emerge in a moment we define $(n \neq 2)$:

$$f^r \equiv (1 - 2/n)^{-1}F^r = (1 - n/2)^{-1}(d^r + g^{pq}C_{pq}{}^r), \qquad f_s = g_{sr}f^r. \tag{2.4$'$}$$

Then the connection in (2.2) in terms of g_{rs} and f_s is

$$\Lambda_{pq}{}^t = C_{pq}{}^t - \tfrac{1}{2}(1 - 2/n)(A^t{}_pf_q + A^t{}_qf_p - g_{pq}f^t). \tag{2.3$'$}$$

The second part of the theorem asserts that we can really associate a *geometry* with the equation; that is, that the geometrical form in which we have cast (2.1) persists unchanged throughout the group of allowable transformations which take (2.1) into equivalent forms. Now this group[1] is the direct product of $\mathfrak{G}_n$: $x^{p'} = f^{p'}(x^q)$, $\det \partial_q x^{p'} \neq 0$: non-singular change of coordinates; $\mathfrak{F}$: multiplication through of the equation by a factor $\tau(x^q) > 0$:

[1] Beside the two groups considered here, Cotton treats a third group $\phi \to \rho\phi$ transforming the unknown. If this further group is adjoined, the equivalence classes will be correspondingly larger.

gauge group. This second part then asserts that the intrinsic geometry is the Weyl geometry W_n (cf. [5], p. 81) defined by the symmetric and linear differential forms g_{rs} and f_s respectively in the precise sense that a) under a transformation of $\mathfrak{G}_n$, (2. 1) goes into $\nabla_{r'}(g^{r's'}\partial_{s'}\phi) = 0$, where the connection $\Lambda_{p'q'}{}^{t'}$ is given by (2. 3)$'$ with

$$\partial_{r'} = \partial/\partial x^{r'}; \qquad g_{r's'} = g_{rs}\partial_{r'}x^r\partial_{s'}x^s, \qquad f_{s'} = f_s\partial_{s'}x^s,$$

and that b) under a transformation of $\mathfrak{F}$ with τ it goes into $'\nabla_r('g^{rs}\partial_s\phi) = 0$, where the connection $'\Lambda_{pq}{}^t$ is given by (2. 3)$'$ with

$$'g_{rs} = \lambda g_{rs}, \qquad 'f_s = f_s + \partial_s \log \lambda; \qquad \lambda = \tau^{-1}.$$

For then g_{rs} and f_s transform under $\mathfrak{G}_n \times \mathfrak{F}$ as in W_n.

Proof. a) is immediate from the invariant form of (2. 2). b) is shown as follows: $g^{rs} \to 'g^{rs} \equiv \tau g^{rs}$, hence $g_{rs} \to 'g_{rs} = \lambda g_{rs}$ for $\lambda = \tau^{-1}$. By (2. 4)$'$,

$$f_s \to 'f_s \equiv (1 - n/2)^{-1}('g_{sr}'d^r + 'g^{pq}'C_{pq}{}^r g_{sr}), \qquad 'd^r \equiv \tau d^r = \lambda^{-1}d^r.$$

This equals

$$f_s + (2-n)^{-1}\lambda^{-1}g^{pq}(2g_{ps}\partial_q\lambda - g_{pq}\partial_s\lambda) = f_s + \partial_s \log \lambda,$$

q. e. d. (It should be noted that the connection (2. 3)$'$ in terms of which (2. 2) is written is *not* the same as the Weyl connection belonging to W_n. In particular, $\Lambda_{pq}{}^r$ is not gauge-invariant.)

It is remarkable that this theory breaks down for $n = 2$. This case is treated briefly at the end of the article.

Of particular interest is the subclass of self-adjoint equations, those which, after multiplication by a suitable (positive) factor, can be written in the form

(2. 1)$'$ $\qquad \partial_r(P^{rs}\partial_s\phi) = 0 \qquad$ [for some set of functions P^{rs}].

This property has a very nice geometrical characterization (remark: this geometrical characterization of self-adjointness is independent of the boundary behavior of ϕ):

THEOREM 2. Equation (2. 1) is self-adjoint if and only if its intrinsic geometry is Riemannian, ($n \neq 2$).

We recall that W_n is Riemannian if and only if $\partial_{[s}f_{r]} = 0$; i. e., there exists a gauge in which $f_r = 0$.

Proof. If W_n is Riemannian, then in the gauge in which $f_r = 0$ by (2.3)′ ∇_r becomes covariant differentiation with respect to $C_{pq}{}^t$. Hence (2.2) reads

$$\nabla_r(g^{rs}\partial_s\phi) \equiv |g|^{-\frac{1}{2}}\partial_r(|g|^{\frac{1}{2}}g^{rs}\partial_s\phi) = 0, \qquad [g \equiv \det g_{rs}].$$

Multiply through by $|g|^{\frac{1}{2}}$; this proves the sufficiency.

Conversely, if equation (2.1) is self-adjoint, then in a suitable gauge $d^r = \partial_s g^{sr}$. From (2.4)′,

$$f_r = (1 - n/2)^{-1} g_{rs}(\partial_p g^{ps} - \partial_p g^{ps} - \tfrac{1}{2} g^{sq}\partial_q \log|g|) = (n-2)^{-1}\partial_r \log|g|,$$

so f_r is a gradient, which proves the necessity.

3. Curvature and the equivalence criterion. Two equations (2.1) are *equivalent* (by this we shall always mean equivalent under the product group $\mathfrak{G}_n \times \mathfrak{F}$) if and only if their intrinsic Weyl geometries are the same. For they both can be written in the form (2.2) with connections of the form (2.3)′ and this prescription is invariant against the transformations considered. It follows that the equivalence problem for these linear equations reduces to the equivalence problem for the Weyl geometry W_n. This closely parallels the usual treatment of the equivalence of quadratic differential forms with slight complications due to the fact that here we have a linear form adjoined and the gauge group as well as coordinate transformation group. It is a noteworthy fact that the equivalence-characterizing system of invariants of the quadratic form and that of the linear form are completely unified in the invariants of W_n. We sketch the proof below, followed by the theorem.

Let g_{pq} and f_r, functions of x^p, define the intrinsic geometry of the first equation, and $'g_{p'q'}$, $'f_{r'}$, functions of $x^{p'}$, that of the second equation. Then we ask whether there exist a gauge transformation $\lambda(x^{q'})$ and coordinate transformation $x^p = f^p(x^{q'})$ such that

(3.1) $\quad 'g_{p'q'} - \lambda g_{pq}A^p{}_{p'}A^q{}_{q'} = 0, \qquad$ (3.2) $\quad 'f_{p'} - f_p A^p{}_{p'} - \Lambda_{p'} = 0,$

where

(3.3) $\quad \partial_{p'}x^p = A^p{}_{p'} \qquad$ (3.3)′ $\quad \partial_{p'}\log\lambda = \Lambda_{p'}.$

Differentiating (3.1) and using (3.2), (3.3), and (3.3)′, we get on rearranging

(3.4) $\quad \Gamma_{p'q'}{}^{r'}A^r{}_{r'} - \Gamma_{pq}{}^r A^p{}_{p'}A^q{}_{q'} - \partial_{p'}A^r{}_{q'} = 0,$

where $\Gamma_{pq}{}^r$ (the linear connection of W_n) is short for

$$\Gamma_{pq}{}^r = C_{pq}{}^r - \tfrac{1}{2}(A^r{}_p f_q + A^r{}_q f_p - g_{pq} f^r)$$

and $\Gamma_{p'q'}{}^{r'}$ is the corresponding expression [2] in $'g_{p'q'}$ and $'f_{r'}$. Differentiating (3.2) and using (3.3) and 3.4), we get on rearranging

$$\nabla_{q'}{}'f_{p'} - \nabla_q f_p A^q{}_{q'} A^p{}_{p'} - \nabla_{q'} \Lambda_{p'} = 0, \tag{3.4$'$}$$

where ∇_q from here on will mean covariant differentiation with respect to $\Gamma_{pq}{}^r$, and $\nabla_{q'}$, correspondingly, with respect to $\Gamma_{p'q'}{}^{r'}$. The problem then reduces to solving the system of partial differential equations (3.3), (3.3)$'$, (3.4), (3.4)$'$ and the finite equations (3.1), (3.2) in the $(n+1)^2$ unknowns x^p, λ, $A^p{}_{p'}$, $\Lambda_{p'}$ as functions of $x^{p'}$.

The integrability conditions of (3.3) are satisfied in virtue of (3.4). The integrability conditions of (3.4) are

$$R_{p'q'r'}{}^{t'} A^t{}_{t'} = R_{pqr}{}^t A^p{}_{p'} A^q{}_{q'} A^r{}_{r'}, \tag{3.5}$$

where $R_{pqr}{}^t$ (the curvature tensor of W_n) stands for

$$R_{pqr}{}^t \equiv -2\partial_{[p}\Gamma_{q]r}{}^t - 2\Gamma_{[p|s|}{}^t \Gamma_{q]r}{}^s$$

correspondingly for $R_{p'q'r'}{}^{t'}$ in terms of $\Gamma_{q'r'}{}^{t'}$, and the infinite sequence of equations obtained by repeated differentiation of (3.5):

$$\nabla_{s'} R_{p'q'r'}{}^{t'} A^t{}_{t'} = \nabla_s R_{pqr}{}^t A^p{}_{p'} A^q{}_{q'} A^r{}_{r'} A^s{}_{s'}, \cdots = \cdots, \text{etc.} \tag{3.6}$$

It is remarkable to note now that the integrability conditions of the other equations, those arising from the linear form, impose no new conditions. For, the integrability conditions of (3.3)$'$ are, from (3.4)$'$,

$$f_{q'p'} = f_{qp} A^q{}_{q'} A^p{}_{p'}, \qquad f_{qp} \equiv \nabla_{[q} f_{p]} = \partial_{[q} f_{p]},$$

correspondingly for $f_{q'p'}$ in terms of $'f_{p'}$, and the infinite sequence of equations arising from these by repeated differentiation. But multiplying (3.5) through by $A^{s'}{}_t$ (the normalized cofactors of $A^t{}_{t'}$), contracting r' and s', and using the identities $R_{pqr}{}^r = nf_{pq}$, $\nabla_s R_{pqr}{}^r = n\nabla_s f_{pq}, \cdots$, we find that these integrability conditions are satisfied in virtue of (3.5), (3.6), $\cdots$. Moreover, the integrability conditions of (3.4)$'$, with the aid of (3.2), come out to be

$$(R_{p'q'r'}{}^{t'} A^t{}_{t'} - R_{pqr}{}^t A^p{}_{p'} A^q{}_{q'} A^r{}_{r'}) f_t = 0$$

and equations arising from these by repeated differentiation. But these are

[2] Note on notation: Γ instead of $'\Gamma$ is written because it is gauge-invariant (cf. the definition equation). The same remark applies to Δ_q, $R_{pqr}{}^t$, and f_{pq}.

satisfied in virtue of (3. 5), (3. 6), $\cdots$. Hence by the well-known theorem (cf. say, [6], Chap. 5, § 7) on systems of partial differential equations the equivalence theorem for the linear equations (2. 1) reads as follows:

THEOREM 3. *Two equations* (2. 1) *whose intrinsic geometries are characterized by the invariants g_{pq}, f_r (functions of x^p) and $'g_{p'q'}$, $'f_{r'}$ (functions of $x^{p'}$) respectively for $n \neq 2$ are equivalent if and only if there exists a positive integer N such that* a) *the sets of equations* (3. 1), (3. 2), *and the first N sets of equations* (3. 5), (3. 6), $\cdots$, *in the unknowns x^p, λ, $A^p{}_{p'}$, $\Lambda_{p'}$ as functions of $x^{p'}$ are compatible and* b) *all sets of solutions of these equations satisfy the* $(N+1)$-th *set of equations.*

Therewith, equations (2. 1) are classified into equivalence classes.

The simplest equation (2. 1) is the ordinary Laplacian equation

$$\delta^{rs}\partial^2{}_{rs}\phi = 0 \qquad [\delta^{rr} = \pm 1; \delta^{rs} = 0, r \neq s]. \tag{3. 7}$$

(The term "Laplacian" here embraces all metric signatures.) Then the equivalence theorem gives us immediately the criterion that any equation (2. 1) be reducible to this form:

COROLLARY. *An equation* (2. 1) *is equivalent to the ordinary Laplacian equation if and only if its intrinsic geometry is flat.*

By *flat* is meant $R_{pqr}{}^t = 0$. (This implies both that the geometry is Riemannian and flat in the Riemannian sense.)

Another application of the complete set of invariants of the intrinsic geometry (in the Riemannian case) is the determination by algebraic means of whether the equation admits any "plane-wave" type solutions. Consider the sets of equations in the unknown $\xi_s(x^p)$

$$(3.\,8) \quad g^{rs}\xi_r\xi_s = 0, \qquad (3.\,9) \quad R_{pqt}{}^s\xi_s = 0,$$

$$(3.\,10) \quad \nabla_l R_{pqt}{}^s\xi_s = 0, \qquad (3.\,11) \quad \nabla^2{}_{ml}R_{pqt}{}^s\xi_s = 0, \cdots.$$

THEOREM 4. *In the Riemannian case $f_{pq} = 0$, there exist solutions of* (2. 1) *of the plane-wave type if and only if there exists a positive integer M such that* a) *the first M of* (3. 9), (3. 10), $\cdots$ *are compatible for the unknown $\xi_s(x^p)$ and* b) *all solutions of these satisfy the* $(M+1)$-th *set of equations, finally,* c) *some solution of these satisfies* (3. 8).

$\phi = F(\int k_s dx^s)$ is defined to be a *plane-wave* if F is any twice differentiable function, $\int$ means the indefinite line integral, and k_s is a null parallel field: $\nabla_r k_s = 0$, $g^{rs}k_r k_s = 0$.

Proof. If and only if such an M exists does there exist a field k_s parallel with respect to the Riemannian space defined by W_n in the gauge for which $f_r = 0$ (cf. [7], § 23); k_s then is a solution of (3.9), (3.10), $\cdots$. If and only if one of these solutions also satisfies (3.8) does there exist a null parallel field. Hence the conditions of the theorem are necessary and sufficient in order that there exist a plane-wave. But every plane-wave yields a solution of (2.1). For in the gauge in which $f_r = 0$ we get from (2.2)

$$g^{rs}\nabla^2{}_{rs}\{F(\textstyle\int k_s dx^s)\} = F''g^{rs}k_r k_s = 0,$$

where F' means the derivative of F with respect to its argument.

4. The case $n = 2$. We add a few words on the anomalous case $n = 2$. The first part of Theorem 1 is still true, but no Weyl geometry W_2 can be associated with the equation. For although g_{pq} and F_r (given by (2.4)) are tensors against $\mathfrak{G}_2$, under $\mathfrak{F}$ they transform as follows:

$$(4.1) \qquad 'g_{pq} = \lambda g_{pq}, \qquad 'F_p = F_p.$$

Hence the geometry is that of a class of conformally related Riemannian spaces V_2, on each of which the same Pfaffian is superimposed.

Theorem 2 holds in the form: Equation (2.1) is self-adjoint if and only if F_r is a gradient.

Of course $\partial_{[s}F_{r]} = 0$ has not now any Riemannian interpretation. In proof, note that if the equation is self-adjoint, then in a coordinate and gauge frame in which it takes the form (2.1)$'$, we have $F_r = \frac{1}{2}\partial_r \log |g|$ (cf. the proof of Theorem 2). Conversely, if F_r is a gradient, $\frac{1}{2}\partial_r \log h$, $h > 0$, say, perform the gauge transformation with $\lambda = h^{\frac{1}{2}} |g|^{-\frac{1}{2}}$ to a frame where $|'g| = h$, $'F_r = F_r = \frac{1}{2}\partial_r \log |'g|$. Then, as in the proof of Theorem 2, in this coordinate and gauge frame $'d^r = \partial_s{}'g^{sr}$ and hence the equation is self-adjoint.

Theorem 3 does not apply. In the present case $n = 2$ the equations we start from in the equivalence problem are

$$(4.2) \qquad 'g_{p'q'} - \lambda g_{pq}A^p{}_{p'}A^q{}_{q'} = 0; \qquad \partial_{p'}x^p = A^p{}_{p'}$$

$$(4.3) \qquad 'F_{p'} - F_p A^p{}_{p'} = 0.$$

Since any two V_2's of the same signature are conformal (cf. [7], § 28), the problem reduces to using the remaining freedom in the coordinate transformations satisfying (4.2) to satisfy (4.3) as well. This was treated at lenth by Cotton ([1], p. 236 *et seq.*) and also elsewhere in the literature.

From the facts that $F_r = 0$ for the ordinary Laplacian equation (3.7), F_r is gauge-invariant for $n = 2$, and any two V_2's of the same signature are conformal, we infer immediately the following

COROLLARY. *The vanishing of F_r is necessary and sufficient in order that the equation* (2.1) *be reducible to ordinary Laplacian form.*

Thus in the case of a definite metric, $F_r = 0$ is the criterion that the solution of (2.1) be reducible to the solution of the Beltrami differential equations.

INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] E. Cotton, "Sur les invariants différentiels de quelques équations linéaires aux dérivées partielles du second ordre," *Annales de l'Ecole Normale Supérieure*, vol. 17 (1900), pp. 211-244.

[2] D. Struik and N. Weiner, "A relativistic theory of quanta," *Journal of Mathematics and Physics*, vol. 7 (1927), pp. 1-23.

[3] T. Levi-Cività, "Sulla trasformazione delle equazioni lineari a derivate parziali del secondo ordine," *Atti del Reale Istituto Veneto*, 3a ser. 15 tomo 72 (1912-13), pp. 1331-1357.

[4] J. Shouten, "La théorie projective de la relativité," *Annales de L'Institut Henri Poincaré*, vol. 5 fasc. 1 (1935), pp. 51-88.

[5] L. Eisenhart, "*Non-Riemannian Geometry*," (1927).

[6] O. Veblen, "*Invariants of Quadratic Differential Forms*," (1927).

[7] L. Eisenhart, "*Riemannian Geometry*," (1926).

KUMMER CONGRUENCES AND THE SCHUR DERIVATIVE.*

By L. CARLITZ.

1. Introduction. Let p be a prime, a an integer not divisible by p, and put

(1.1) $$\Delta a^{p^m} = (a^{p^{m+1}} - a^{p^m})/p^{m+1}, \qquad \Delta^r a^{p^m} = \Delta(\Delta^{r-1} a^{p^m}) \quad (r = 2, 3, \cdots).$$

Schur [6] proved that the derivatives $\Delta a^{p^m}, \Delta^2 a^{p^m}, \cdots, \Delta^{p-1} a^{p^m}$ are all integral. If $a^{p-1} \equiv 1 \pmod{p^2}$ then all the derivatives $\Delta^r a^{p^m}$ are integral, while if $a^{p-1} \not\equiv 1 \pmod{p^2}$, then every number $\Delta^p a^{p^m}$ has exactly the denominator p. A. Brauer [1] gave another proof of these results; about the same time Zorn [7] also proved these and other results by p-adic methods.

In the present paper we consider sequences of rational numbers $\{a_m\}$ that are integral $\pmod{p}$, where p is a fixed prime. (More generally the a_m may be integral p-adic numbers.) Now suppose that the a_m satisfy

(1.2) $$\sum_{s=0}^{r} (-1)^{r-s} C_s^r a_p^{r-s} a_{m+s(p-1)} \equiv 0 \pmod{p^r}$$

for $m \geqq r \geqq 1$. We shall call (1.2) Kummer's congruence for $\{a_m\}$. The same term may also be used for the stronger congruence

(1.3) $$\sum_{s=0}^{r} (-1)^{r-s} C_s^r a_p^{(r-s)b/(p-1)} a_{m+sb} \equiv 0 \pmod{p^{re}},$$

where $p^{e-1}(p-1) \mid b$ and $m \geqq re$, $e \geqq 1$, $r \geqq 1$. As we shall see, (1.2) implies (1.3).

Next, generalizing (1.1), we define

(1.4) $$\Delta a_{p^m} = (a_{p^{m+1}} - a_p^{p^m} a_{p^m})/p^{m+1},$$
$$\Delta^r a_{p^m} = (\Delta^{r-1} a_{p^{m+1}} - a_p^{p^{m+r-1}} \Delta^{r-1} a_{p^m})/p^{m+1}.$$

Then we shall show that if the sequence $\{a_m\}$ satisfies (1.2) it follows that the numbers

(1.5) $$\Delta a_{p^m}, \Delta^2 a_{p^m}, \cdots, \Delta^{p-1} a_{p^m}$$

are integral for all m. Moreover the residues of these numbers $\pmod{p^m}$ are specified. The method of proof is essentially that used in [4]. We also

* Received October 16, 1952.

consider the vase in which the a_m are polynomials in an indeterminate u with integral coefficients.

We also discuss some applications (§§ 6, 7) of these results as well as a generalization (§ 8).

2. Kummer's congruences. To show that (1.2) implies (1.3) we remark first of all that by the binomial expansion

$$(2.1) \qquad t^{p(p-1)} - u^p = (t^{p-1} - u)^p + p(t^{p-1} - u)f(t, u),$$

where $f(t, u)$ denotes a polynomial in the indeterminates t, u with integral coefficients. Then by a straightforward induction (2.1) yields the more general formula

$$(2.2) \qquad t^{p^e(p-1)} - u^{p^e} = (t^{p-1} - u)^{p^e} + \sum_{i=1}^{e} p^i (t^{p-1} - u)^{p^{e-i}} f_i(t, u)$$

for all $e \geqq 1$; the $f_i(t, u)$ are polynomials with integral coefficients. We shall also require the formula obtained from (2.2) by raising both members to the r-th power and expanding the right side; this may be referred to as $(2.2)_r$.

Now returning to (1.2) it is clear that the left side may be written in the following form:

$$(2.3) \qquad (E^{p-1} - a_p)^r a_m \qquad (Ea_m = a_{m+1}),$$

provided we agree that E operates only on m (and therefore commutes with a_p). Similarly the left side of (1.3) may be written

$$(2.4) \qquad (E^b - a_p{}^{b/(p-1)})^r a_m.$$

Let us consider first the case $e = 1$, so that $b = c(p-1)$, $c \geqq 1$. Comparing (2.3) with (2.4) it is clear since $E^{c(p-1)} - a_p{}^c$ contains $E^{p-1} - a_p$ as a factor that (1.3) holds in this case. For arbitrary $e \geqq 1$, we take $t = E^{c(p-1)}$, $u = a_p{}^c$ in $(2.2)_r$ and apply the special case of (1.3) just obtained. Each term in the right member is seen to be divisible by at least p^{er}. This proves

THEOREM 1. *If* (1.2) *holds for* $m \geqq r \geqq 1$ *then* (1.3) *also holds for all* $m \geqq er$, $r \geqq 1$, $e \geqq 1$.

For some applications we shall be interested in the following extension. Let $a_m = a_m(u)$ denote a polynomial in an indeterminate u with rational coefficients that are integral (mod p). (If $f(u)$ is such a polynomial, then the statement $f(u) \equiv 0 \pmod{p^r}$ means that each coefficient of $f(u) \equiv 0 \pmod{p^r}$.) It is then clear what meaning is to be attached to (1.2) and

(1.3) in this situation. Moreover it is evident that the above proof applies. We may state

THEOREM 1′. *If the a_m are polynomials in u with integral coefficients, then Theorem 1 holds.*

We also remark that if in (1.2) we replace a_p by a number congruent to it (mod p), then the congruence will continue to hold; the same is therefore true of (1.3). Indeed we have the identity

$$\sum_{s=0}^{r} (-1)^{r-s} C_s{}^r a_{m+s(p-1)} (a_p + t)^{r-s} = \sum_{i=0}^{r} (-1)^i C_i{}^r t^i \sum_{s=0}^{r-i} (-1)^s C_s{}^{r-i} a_{m+s(p-1)} a_p{}^{r-i-s},$$

from which the above statement is evident.

3. Some lemmas. We shall require the following lemmas.

LEMMA 1.

$$\prod_{i=0}^{r-1} (x - p^i y) = \sum_{i=0}^{r} (-1)^i [r, i] p^{\frac{1}{2}i(i-1)} x^{r-i} y^i, \tag{3.1}$$

where $[r, i] = \dfrac{(p^r - 1) \cdots (p^{r-i+1} - 1)}{(p - 1) \cdots (p^i - 1)} = [r, r-1],\ [r, 0] = 1.$

LEMMA 2. *If* $e_i = (p^i - 1)/(p - 1)$, $e_0 = 0$, *then*

$$\Delta^r a_{p^m} = p^{-rm-\frac{1}{2}r(r+1)} \sum_{i=0}^{r} (-1)^i [r, i] p^{\frac{1}{2}i(i-1)} a_{p^{m+r-i}} a_p{}^{p^{m+r-i} e_i}. \tag{3.2}$$

LEMMA 3. *Let*

$$W_{k,r} = \sum_{i=0}^{k} (-1)^{k-i} [k, i] C_r{}^{e_i} p^{\frac{1}{2}(k-i)(k-i-1)},$$

where $e_i = (p^i - 1)/(p - 1)$ *and* $C_r{}^m$ *denotes a binomial coefficient. Then*

$$W_{k,r} = 0,\ p^{\frac{1}{2}r(r-1)} \prod_{i=1}^{r} e_i / r!,\ p^{\frac{1}{2}k(k-1)} U_{k,r} / r! \tag{3.3}$$

according as $r < k$, $r = k$, $r > k$, *respectively, where* $U_{k,r}$ *is an integer.*

Lemma 1 is familiar. Lemma 2 is a slight extension of a formula of Schur. For Lemma 3, compare [4], Lemma 2.

4. A formula for $\Delta^r a_{p^m}$. If in (1.3) we take $b = p^e(p-1)$, $r = 1$, $m = p^e$, we get $a_{p^{e+1}} - a_p{}^{p^e} a_{p^e} \equiv 0 \pmod{p^{e+1}}$. In other words, using the

first of (1.4), we see that Δa_{p^e} is integral. In order to treat the general case we replace b and m in (1.3) by $p^m(p-1)$ and p^m, respectively, and put

$$\sum_{s=0}^{r} (-1)^{r-s} C_s^r a_p^{(r-s)p^m} a_{p^m(1+s(p-1))} = p^{r(m+1)} Q_{m,r}, \tag{4.1}$$

so that $Q_{m,r}$ is integral (mod p). Now it is easily verified that (4.1) implies

$$a_{p^m(1+r(p-1))} = \sum_{s=0}^{r} C_s^r a_p^{(r-s)p^m} p^{s(m+1)} Q_{m,s}. \tag{4.2}$$

In particular for $r = e_k = (p^k-1)/(p-1)$, (4.2) becomes

$$a_{p^{m+k}} = \sum_{s=0}^{e_k} C_s^{e_k} a_p^{(e_k-s)p^m} p^{s(m+1)} Q_{m,s}. \tag{4.3}$$

Substituting from (4.3) in (3.2) we get

$$\begin{aligned} p^{rm+\frac{1}{2}r(r+1)}\Delta^r a_{p^m} &= \sum_{i=0}^{r} (-1)^{r-i} p^{\frac{1}{2}(r-i)(r-i-1)} a_p^{p^{m+i}e_{r-i}} [r, i] \\ &\qquad \cdot \sum_{s=0}^{e_i} C_s^{e_i} a_p^{(e_i-s)p^m} p^{s(m+1)} Q_{m,s} \\ &= \sum_{s=0}^{e_r} p^{s(m+1)} a_p^{(e_r-s)p^m} Q_{m,s} \sum_{i=0}^{r} (-1)^{r-i} [r, i]\, C_s^{e_i} p^{\frac{1}{2}(r-i)(r-i-1)} \\ &= \sum_{s=0}^{e_r} p^{s(m+1)} a_p^{(e_r-s)p^m} Q_{m,s} W_{r,s} \\ &= (1/r!)\, p^{r(m+1)+\frac{1}{2}r(r-1)} a_p^{(e_r-r)p^m} Q_{m,r} \prod_{i=1}^{r} e_i \\ &\qquad + \sum_{s=r+1}^{e_r} (1/s!)\, p^{s(m+1)+\frac{1}{2}r(r-1)} a_p^{(e_r-s)p^m} Q_{m,s} U_{m,s}, \end{aligned}$$

by (3.3); $W_{m,s}$ and $U_{m,s}$ have the same meaning as in Lemma 3. Thus it follows that

$$\begin{aligned} \Delta^r a_{p^m} = (1/r!)\, a_p^{(e_r-r)p^m} Q_{m,r} \prod_{i=1}^{r} e_i \\ + \sum_{s=r+1}^{e_r} (1/s!)\, p^{(m+1)(s-r)} a_p^{(e_r-s)p^m} Q_{m,s} U_{m,s}. \end{aligned} \tag{4.4}$$

It may be of interest to mention a variant of (4.4). In (1.3) we replace both b and m by $p^m(p-1)$ and put

$$\sum_{s=0}^{r} (-1)^{r-s} C_s^r a_p^{(r-s)p^m} a_{p^m(s+1)(p-1)} = p^{r(m+1)} Q'_{m,r}, \tag{4.5}$$

so that $Q'_{m,r}$ is integral $(\mathrm{mod}\ p)$. Then (4. 5) implies

$$a_{p^m(r+1)(p-1)} = \sum_{s=0}^{r} C_s{}^r a_p{}^{(r-s)p^m} p^{s(m+1)} Q'_{m,s}. \tag{4. 6}$$

In particular for $r+1 = p^k$, (4. 6) becomes

$$a_{(p-1)p^{m+k}} = \sum_{s=0}^{p^k-1} C_s{}^{p^k-1} a_p{}^{(p^k-1-s)p^m} p^{s(m+1)} Q'_{m,s}. \tag{4. 7}$$

Substituting from (4. 7) in (1. 3) we get, very much as before,

$$\begin{aligned} \Delta^r a_{(p-1)p^m} = {} & (1/r!)\, a_p{}^{(p^r-1-r)p^m} Q'_{m,r} \prod_{i=1}^{r} (p^i - 1) \\ & + \sum_{s=r+1}^{p^r-1} (1/s!)\, p^{(m+1)(s-r)} a_p{}^{(p^r-1-s)p^m} Q'_{m,s} U'_{m,s}, \end{aligned} \tag{4. 8}$$

where $U'_{m,s}$ is integral and

$$\Delta a_{(p-1)p^m} = (a_{(p-1)p^{m+1}} - a_p{}^{(p-1)p^m} a_{(p-1)p^m}) / p^{m+1},$$

and

$$\Delta^r a_{(p-1)p^m} = (\Delta^{r-1} a_{(p-1)p^{m+1}} - a_p{}^{(p-1)p^{m+r-1}} \Delta^{r-1} a_{(p-1)p^m}) / p^{m+1}.$$

Additional formulas of this kind are easily obtained.

5. The main results. By means of (4. 4) it is easy to derive the main results concerning $\Delta^r a_{p^m}$ (compare [4], § 3). It is evidently only necessary to examine $p^{(m+1)(s-r)}/s!$ $(s > r)$. Let $r \leqq p$; then $p^{s-r}/s!$ is integral $(\mathrm{mod}\ p)$ and is indeed divisible by p unless (i) $s = p$, $r = p-1$, or (ii) $s = p+1$, $r = p$. We may now state the following two theorems.

THEOREM 2. *$\Delta^r a_{p^m}$ is integral for $1 \leqq r \leqq p-1$, while $\Delta^p a_{p^m}$ has the denominator p provided $Q_{m,p} \not\equiv 0$ $(\mathrm{mod}\ p)$.*

THEOREM 3. *For $1 \leqq r \leqq p$,*

$$\Delta^r a_{p^m} \equiv a_p{}^{(e_r-r)p^m} \frac{Q_{m,r} \prod_{i=1}^{r} (p^i - 1)}{r!(p-1)^r} \qquad (\mathrm{mod}\ p^m), \tag{5. 1}$$

where $Q_{m,r}$ is defined by (4. 1).

Similarly using (4. 8) we get, for $1 \leqq r \leqq p$,

$$\Delta^r a_{(p-1)p^m} \equiv (1/r!)\, a_p{}^{(p^r-1-r)p^m} Q'_{m,r} \prod_{i=1}^{r} (p^i - 1) \qquad (\mathrm{mod}\ p^m), \tag{5. 2}$$

where $Q'_{m,r}$ is defined by (4. 5); for $r < p-1$, (5. 2) holds $(\mathrm{mod}\ p^{m+1})$.

In the next place, if as in Theorem 1′ we suppose the a_m polynomials

in u with integral coefficients then Theorems 2 and 3 will continue to hold. The corresponding results will be referred to as Theorem 2′ and Theorem 3′. It seems unnecessary to state these theorems explicitly.

6. Applications to the coefficients of the Jacobi ellipic functions. We now assume $p > 2$. The writer has proved [2] that if $l = k^2$ is rational $(\mathrm{mod}\, p)$, and

$$sn\, x = sn(x, k^2) = \sum_{m=1}^{\infty} a_m x^m / m\,!, \tag{6.1}$$

then the coefficients a_m satisfy (1.2). Consequently we have at once

THEOREM 4. *The coefficients a_m defined by* (6.1) *satisfy Theorems* 2 *and* 3.

However it is evident that we can say a great deal more. For not only does a_m satisfy (1.3) but so also does a_{m+t} and a_{mt}. Accordingly we state

THEOREM 5. $\Delta^r a_{p^m t}$ *and* $\Delta^r a_{p^m+t}$ *are integral* $(\mathrm{mod}\, p)$, *where t is an arbitrary integer.*

It is also easily seen that the last two theorems hold for the coefficients b_m defined by

$$sn^h x\, cn^{h'} x\, dn^{h''} x = \sum_{m=h}^{\infty} b_m x^m / m\,!, \tag{6.2}$$

where h, h', h'' are integers and $h \geqq 0$.

As for the coefficients β_m defined by

$$x / sn\, x = \sum_{m=0}^{\infty} \beta_m x^m / m\,! \tag{6.3}$$

we have by [2], Theorem 4

$$\sum_{i=0}^{r} (-1)^i C_i^r a_p^{r-i} \tau_{m+i(p-1)} \equiv 0 \qquad (\mathrm{mod}\, p^r), \tag{6.4}$$

where $\tau_m = \beta_m / m$, $p - 1 \nmid m$ and $m > r \geqq 1$. Comparison of (6.4) with (1.2) shows that the latter is not quite satisfied in this case. However it is easily verified that the following theorem holds.

THEOREM 6. *Let $\tau_m = \beta_m / m$, where β_m is defined by* (6.3). *Then*

$$\Delta^r \tau_{k+p^m} \qquad (r < p, r \leqq m, k \geqq 1) \tag{6.5}$$

is integral provided $p - 1 \nmid k + 1$.

In the next place if in (6.1) we use the fuller notation

$$(6.6) \qquad sn(x, u) = \sum_{m=1}^{\infty} A_m(u) x^m / m\,!,$$

where $A_m(u)$ is a polynomial in the indeterminate u with integral coefficients. Then by [3], Theorem 3, we see that $A_m(u)$ also satisfies the congruence (1.2). Thus Theorem 1′ as well as the later discussion applies and we obtain

THEOREM 7. *The coefficients $A_m(u)$ defined by (6.4) satisfy Theorems 2′ and 3′.*

It is clear how the results concerning (6.2) and (6.5) can be carried over to the present situation; we shall accordingly not take the space to state these theorems explicitly.

7. Application to Eulerian polynomials. For a second application we consider the Eulerian polynomials $A_k(u)$ which may be defined by means of $A_k(u) = (u-1)^k H_k(u)$, where

$$(1-u)/(e^x - u) = \sum_{k=0}^{\infty} H_k(u) x^k / k\,!.$$

Then Frobenius [5] has proved that $H_k(u)$ satisfies a congruence of the form (1.3), from which it follows that the polynomial $A_k(u)$ also satisfies a congruence of the same sort. Consequently applying Theorem 2′ we can assert that the polynomials

$$(7.1) \qquad \Delta^r A_{kp^m}(u), \qquad \Delta^r A_{k+p^m}(u)$$

have integral coefficients for $r < p$. We can also specify the residues of (7.1) (mod p^m) in terms of the corresponding Kummer quotients.

8. A generalization. As in [4, § 5] the definition (1.4) can be generalized. Let us define

$$(8.1) \qquad \begin{aligned} \Delta_p a_{mp^i} &= (a_{mp^{i+1}} - a_p{}^{mp^i} a_{mp^i})/p^{i+1}, \\ \Delta_p{}^r a_{mp^i} &= (\Delta_p{}^{r-1} a_{mp^{i+1}} - a_p{}^{mp^{i+r-1}} \Delta_p{}^{r-1} a_{mp^i})/p^{i+1}. \end{aligned}$$

Then it is clear that $\Delta_p{}^r \Delta_q{}^s = \Delta_q{}^s \Delta_p{}^r$. Now put

$$(8.2) \qquad \delta_k{}^{(r)} a_{mk} = \Delta_{p_1}{}^{r_1} \cdots \Delta_{p_s}{}^{r_s} a_{mk},$$

where $k = p_1{}^{e_1} \cdots p_s{}^{e_s}$. Theorem 2 implies the following generalization.

THEOREM 8. *Let k be a fixed integer $\geqq 2$. If (1.2) holds for all primes p_i dividing k and $r_i < p_i$, $j = 1, \cdots, s$, then $\delta_k a_k$ is integral.*

In particular when all the r_i are equal we may write $\delta_k{}^r$ in place of $\delta_k{}^{(r)}$ in (8.2); it is possible to set up explicit formulas for both $\delta_k{}^r$ and $\delta_k{}^{(r)}$ (compare [4], § 5).

DUKE UNIVERSITY.

REFERENCES.

[1] A. Brauer, "Über eine Erweiterung des kleinen Fermatschen Satzes," *Mathematische Zeitschrift*, vol. 42 (1937), pp. 255-262.

[2] L. Carlitz, "Congruences for the coefficients of the Jacobi elliptic functions," *Duke Mathematical Journal*, vol. 16 (1949), pp. 297-302.

[3] ———, "Congruences connected with the power series expansions of the Jacobi elliptic functions," *Duke Mathematical Journal*, vol. 20 (1953), pp. 1-12.

[4] ———, "Some theorems on the Schur derivative," *Pacific Journal of Mathematics*, vol. 3 (1953), pp. 321-332.

[5] G. Frobenius, "Über die Bernoullischen Zahlen und die Eulerschen Polynome," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1910), pp. 809-847.

[6] I. Schur, "Ein Beitrag zur elementaren Zahlentheorie," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1933), pp. 145-151.

[7] M. Zorn, "p-adic analysis and elementary number theory," *Annals of Mathematics* (2), vol. 38 (1937), pp. 451-464.

SOME CONGRUENCES OF VANDIVER.*

By L. Carlitz.

1. Introduction. Professor Vandiver has kindly called the writer's attention to the following congruence for the Bernoulli numbers:

$$h_1^{n_1}\cdots h_k^{n_k}(\lambda_1 h_1^{p-1}+\cdots+\lambda_k h_k^{p-1})^r \equiv 0 \qquad (\operatorname{mod} p^r, p^{n_1-1},\cdots, p^{n_k-1}) \tag{1.1}$$

$$(n_i \not\equiv 0 \ (\operatorname{mod} p-1), i=1,\cdots,k),$$

where the left-member is expanded in full and B_m/m substituted for h_i^m in the result; B_m is the Bernoulli number in the even suffix notation $((B+1)^m = B_m, m>1)$, the λ's are rational integers such that

$$\lambda_1+\cdots+\lambda_k \equiv 0 \qquad (\operatorname{mod} p), \tag{1.2}$$

and p is an odd prime. For example (1.1) implies in particular

$$\sum_{s=0}^{r}(-1)^s C_s^r \frac{B_{m+(r-s)(p-1)}}{m+(r-s)(p-1)}\,\frac{B_{n+s(p-1)}}{n+s(p-1)} \equiv 0 \qquad (\operatorname{mod} p^r)$$

provided $p-1\nmid m$, $p-1\nmid n$, $m>r$, $n>r$. The congruence (1.1) is derived from a more general result proved in [5].

In the present note we wish to point out a generalization of (1.1) that is valid for certain sequences including the Bernoulli and Euler numbers. Let $\{a_m\}$ be a sequence of rational numbers that are integral $(\operatorname{mod} p)$ and assume that

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r a_{m+s(p-1)}a'^{\,r-s}_p \equiv 0 \qquad (\operatorname{mod} p^r, p^m), \tag{1.3}$$

where a'_p is integral $(\operatorname{mod} p)$. We may call (1.3) Kummer's congruence for $\{a_m\}$. Now let $\{b_m\}$ denote another sequence that satisfies

$$\sum_{s=0}^{r}(-1)^{r-s}C_s^r b_{m+s(p-1)}b'^{\,r-s}_p \equiv 0 \qquad (\operatorname{mod} p^r, p^m),$$

where b'_p is integral $(\operatorname{mod} p)$. Then we prove that

$$\sum_{s=0}^{r}C_s^r a_{m+(r-s)(p-1)}b_{n+s(p-1)}(\mu a'_p)^s(\lambda b'_p)^{r-s} \equiv 0 \qquad (\operatorname{mod} p^r, p^m, p^n), \tag{1.4}$$

* Received January 16, 1953.

provided $\lambda + \mu \equiv 0 \pmod{p}$. The left member of (1.4) can be written briefly as

$$a^m b^n (\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p)^r. \tag{1.5}$$

More generally we prove that if $p^{e-1}(p-1) \mid u$, $p^{e-1}(p-1) \mid v$ and $\lambda_1 + \cdots + \lambda_k \equiv 0 \pmod{p^e}$, then

$$a^m b^n (\lambda a^u b'_p{}^{v/(p-1)} + \mu b^v a'_p{}^{u/(p-1)})^r \equiv 0 \qquad (\operatorname{mod} p^{re}, p^m, p^n), \tag{1.6}$$

where it is understood that the left member of (1.6) is to be expanded in full and $a^{m+(r-s)u}$, b^{n+sv} replaced by $a_{m+(r-s)u}$, b_{n+sv}, respectively. For the generalization of (1.6) corresponding to (1.1), see (3.5) below.

We also indicate a few applications of these results, particularly to the coefficients of the Jacobi elliptic functions, in § 4.

For some related results on sequences satisfying Kummer's congruences, see [3].

2. Proof of (1.4) and (1.6). To prove (1.4) we require the following identity

$$a^m b^n (\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p)^r = \sum_{i+j \leqq r} \frac{r!}{i!\, j!\, (r-i-j)!} \text{ times} \tag{2.1}$$

$$a^m (a^{p-1} - a'_p)^i b^n (b^{p-1} - b'_p)^j (\lambda b'_p)^i \cdot (\mu a'_p)^j (\lambda + \mu)^{r-i-j} (a'_p b'_p)^{r-i-j}.$$

Indeed (2.1) is almost immediate if we raise both members of the equation

$$\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p = (a^{p-1} - a'_p) \lambda b'_p + \mu (b^{p-1} - b'_p) a'_p + (\lambda + \mu) a'_p b'_p$$

to the r-th power and subsequently multiply by $a^m b^n$. If we prefer, we can give a straightforward proof of (2.1) without the use of symbolic notation.

Now assume that $\{a_m\}$ satisfies (1.3) and that $\{b_m\}$ satisfies a like congruence; also that $\lambda + \mu \equiv 0 \pmod{p}$. Then it is clear that each term in the right member of (2.1) is $\equiv 0$ $(\operatorname{mod} p^i \cdot p^j \cdot p^{r-i-j}, p^m, p^n)$. Thus it follows that

$$a^m b^n (\lambda a^{p-1} b'_p + \mu b^{p-1} a'_p)^r \equiv 0 \qquad (\operatorname{mod} p^r, p^m, p^n), \tag{2.2}$$

so that we have proved (1.4).

In the next place we remark that if the sequence $\{a_m\}$ satisfies (1.3) then indeed the following congruence holds:

$$a^m (a^w - a'_p{}^{w/(p-1)})^r \equiv 0 \qquad (\operatorname{mod} p^{re}, p^m), \tag{2.3}$$

where $p^{e-1}(p-1) \mid w$. (For proof see [2], Theorem 1; while in that theorem

a'_p was taken equal to a_p, this is not required in the proof.) In place of (2.1) we now employ the identity

$$(2.4)\qquad a^m b^n (\lambda a^u b'_p{}^{v/(p-1)} + \mu b^v a'_p{}^{u/(p-1)})^r$$
$$= \sum_{i+j\leq r} \frac{r!}{i!\,j!\,(r-i-j)!}\, a^m (a^u - a'_p{}^{u/(p-1)})^i b^n (b^v - b'_p{}^{v/(p-1)})^j$$
$$\times (\lambda b'_p{}^{v/(p-1)})^i (\mu a'_p{}^{u/(p-1)})^j (\lambda+\mu)^{r-i-j} (a'_p{}^{u/(p-1)} b'_p{}^{v/(p-1)})^{r-i-j}.$$

As above, (2.4) follows readily from

$$\lambda a^u b'_p{}^{v/(p-1)} + \mu b^v a'_p{}^{u/(p-1)} = (a^u - a'_p{}^{u/(p-1)})\lambda b'_p{}^{v/(p-1)}$$
$$+ (b^v - b'_p{}^{v/(p-1)})\mu a'_p{}^{u/(p-1)} + (\lambda+\mu) a'_p{}^{u/(p-1)} b'_p{}^{v/(p-1)}.$$

Hence with the assumptions used in proving (2.2) we have first that $\{a_m\}$ satisfies (2.3) and that $\{b_m\}$ satisfies a like congruence. Assuming also that $\lambda + \mu \equiv 0 \pmod{p^e}$ it is clear that (2.4) implies

$$a^m b^n (\lambda a^u b'_p{}^{v/(p-1)} + \mu b^v a'_p{}^{u/(p-1)})^r \equiv 0 \qquad (\mathrm{mod}\ p^{re}, p^m, p^n),$$

so that we have proved (1.6).

Note that (1.6) continues to hold if we replace a'_p, b'_p by a''_p, b''_p, respectively, where $a'_p \equiv a''_p$, $b'_p \equiv b''_p \pmod{p}$. This is a fairly easy consequence of the fact that $a'_p{}^{p^{e-1}} \equiv a''_p{}^{p^{e-1}}$, $b'_p{}^{p^{e-1}} \equiv b''_p{}^{p^{e-1}} \pmod{p^e}$, so that the left member of (1.6) is unchanged $\pmod{p^e}$ by the replacement.

3. The general case. The statement of the general case corresponding to (1.1) and (1.6) is somewhat complicated notationally. Let $\{a_{i,m}\}$, $i = 1, \cdots, k$, denote k sequences each satisfying

$$(3.1)\qquad a_i^m (a_i^{p-1} - a'_{i,p})^r \equiv 0 \qquad (\mathrm{mod}\ p^r, p^m),$$

where after expansion of the left member $a_i^{m+s(p-1)}$ is replaced by $a_{i,m+s(p-1)}$. Let $\lambda_1, \cdots, \lambda_k$ denote rational numbers that are integral $\pmod{p}$ and satisfy

$$(3.2)\qquad \lambda_1, \cdots, \lambda_k \equiv 0 \qquad (\mathrm{mod}\ p^e).$$

Let $u_1, \cdots, u_k$ denote integers such that

$$(3.3)\qquad p^{e-1}(p-1) \mid u_i \qquad (i = 1, \cdots, k).$$

Also put

$$(3.4)\qquad c_{i,p} = \prod_{j=1}^{k}{}' a'_{i,p}{}^{u_i/(p-1)}, \text{ where } j \neq i.$$

We now state the following

THEOREM 1. *If* (3.1), (3.2), (3.3) *and* (3.4) *hold, then*

$$a_1^{m_1}\cdots a_k^{m_k}(\lambda_1 a_1^{u_1}c_{1,p}+\cdots+\lambda_k a_k^{u_k}c_{k,p})^r\equiv 0 \pmod{p^{re}, p^{m_1},\cdots,p^{m_k}}, \tag{3.5}$$

where after expansion of the left member $a_i^{m_i+su_i}$ *is replaced by* a_{i,m_i+su_i}.

The proof of (3.5) is very much like the proof of (1.6) except that (2.4) is replaced by a somewhat more elaborate formula, the basis of which is the evident identity

$$\sum_{i=1}^{k}\lambda_i a_i^{u_i}c_{i,p}=\sum_{i=1}^{k}\lambda_i(a_i^{u_i}-a'^{\,u_i/(p-1)}_{i,p})c_{i,p}+\sum_{i=1}^{k}\lambda_i\prod_{j=1}^{k}a'^{\,u_i/(p-1)}_{i,p}.$$

In view of the complicated nature of the hypotheses leading to (3.5) as well as of the formula itself, it may be of interest to state explicitly the following special case which indeed corresponds more closely to (1.1). We assume (3.1) but in place of (3.2), (3.3), (3.4) we suppose that

$$\lambda_1+\cdots+\lambda_k\equiv 0 \pmod{p}, \tag{3.2$'$}$$

$$c_{i,p}=\prod_{j=1}^{k}{}' a'_{i,p}, \text{ where } j\neq i. \tag{3.4$'$}$$

Then we can assert that if the k sequences $\{a_{i,m}\}$ satisfy (3.1), (3.2)$'$ and (3.4)$'$, it follows that

$$a_1^{m_1}\cdots a_k^{m_k}(\lambda_1 a_1^{p-1}c_{1,p}+\cdots+\lambda_k a_k^{p-1}c_{k,p})^r\equiv 0 \pmod{p^{r}, p^{m_1},\cdots,p^{m_k}}. \tag{3.5$'$}$$

As a corollary of Theorem 1 we state

THEOREM 2. *Assume in addition to the hypothesis of Theorem* 1 *that* $a'_{1,p}\equiv\cdots\equiv a'_{k,p} \pmod{p}$. *Then we have*

$$a_1^{m_1}\cdots a_k^{m_k}(\lambda_1 a_1^{u_1}+\cdots+\lambda_k a_k^{u_k})^r\equiv 0 \pmod{p^{re}, p^{m_1},\cdots,p^{m_k}}. \tag{3.6}$$

In particular Theorem 2 applies when the sequences $\{a_{i,m}\}$ are identical.

To prove the theorem we note first that (3.6) is an immediate consequence of (3.5) when $p\nmid a'_p$. On the other hand when $p\mid a'_p$, it follows readily from (2.3) and the identity

$$a^{m+sw}=\sum_{t=0}^{s}C_t{}^s a^m(a^w-a'^{\,w/(p-1)}_p)^{s-t}a'^{\,tw/(p-1)}_p$$

that $a_{m+sw}\equiv 0 \pmod{p^{se}}$ and this in turn leads to (3.6) in the case $p\mid a'_p$.

4. Some applications. In the first place it is clear that (1.6) and (3.5) hold for the Euler numbers E_m (in the even suffix notation) and for the numbers $h_m = B_m/m$ occurring in (1.1) as well as for various related sequences. Moreover these numbers can be combined in various ways. For example we may mention the special case

$$h^m E^n (h^{p-1} - E^{p-1})^r \equiv 0 \qquad (\mathrm{mod}\ p^r, p^{m-1}, p^n), \tag{4.1}$$

provided $p > 2$ and $(p-1) \nmid m$. Additional results of this kind can be stated without any difficulty.

It is perhaps of greater interest to point out one or two sequences of a more recondite sort that also satisfy the hypotheses of our theorems. Let $p > 2$ and let l be a rational number that is integral $(\mathrm{mod}\ p)$. Let $sn\, x = sn(x, l)$ denote the Jacobi elliptic function with modulus $k^2 = l$. Put

$$sn\, x = \sum_0^\infty A_{2m+1} x^{2m+1}/(2m+1)! \qquad (c_1 = 1) \tag{4.2}$$

and

$$x/sn\, x = \sum_0^\infty \beta_{2m} x^{2m}/(2m)! \qquad (\beta_0 = 1). \tag{4.3}$$

Then [1] the sequence $\{A_m\}$ satisfies $A^m(A^{p-1} - A_p)^r \equiv 0 \ (\mathrm{mod}\ p^r, p^m)$; also the sequence $\{\eta_m\} = \{\beta_m/m\}$ satisfies $\eta^m(\eta^{p-1} - A_p)^r \equiv 0 \ (\mathrm{mod}\ p^r, p^{m-1})$ provided $(p-1) \nmid m$. Thus our results hold for the sequences $\{A_m\}$ and $\{\eta_m\}$. In particular we can state such congruences as

$$A_1^{m_1} \cdots A_k^{m_k} (\lambda_1 A_1 + \cdots + \lambda_k A_k)^r \equiv 0 \qquad (\mathrm{mod}\ p^r, p^{m_1}, \cdots, p^{m_k}) \tag{4.4}$$

and

$$\eta_1^{m_1} \cdots \eta_k^{m_k} (\lambda_1 \eta_1 + \cdots + \lambda_k \eta_k)^r \equiv 0 \qquad (\mathrm{mod}\ p^r, p^{m_1-1}, \cdots, p^{m_k-1}), \tag{4.5}$$

where in (4.5) it is assumed that $(p-1) \nmid m_i$. Also corresponding to (4.1) we can state congruences of the type

$$A^m \eta^n (A^{p-1} - \eta^{p-1})^r \equiv 0 \qquad (\mathrm{mod}\ p^r, p^m, p^{n-1}), \tag{4.6}$$

where $(p-1) \nmid n$. In the second place we may vary the modulus l. We accordingly consider k sequences $\{A_m(l_i)\}$ defined by means of

$$sn(x, l_i) = \sum_0^\infty A_{2m+1}(l_i) x^{2m+1}/(2m+1)! \qquad (i = 1, \cdots, k);$$

we define $\beta_m(l_i)$ in the obvious way. Then Theorem 1 yields

$$A^{m_1}(l_i)\cdots A^{m_k}(l_k)(\lambda_1 A^{u_1}(l_i)C(l_1)+\cdots+\lambda_k A^{u_k}(l_k)C(l_k))^r\equiv 0$$

(mod $p^{re}, p^{m_1}, \cdots, p^{m_k}$), where the u_i satisfy (3.3) and as in (3.4) $C(l_i)=\prod_{j=1}^{k}{}' A^{u_i/(p-1)}(l_i)$, where $j\neq i$. Similarly

$$\eta_1{}^{m_1}(l_1)\cdots\eta^{m_k}(l_k)(\lambda_1\eta^{u_1}(l_1)C(l_1)+\cdots+\lambda_k\eta^{u_k}(l_k)C(l_k))^r\equiv 0$$

(mod $p^{re}, p^{m_1-1}, \cdots, p^{m_k-1}$), where now $(p-1)\nmid m_i$.

As a special case of these results concerning the coefficients of the Jacobi functions we may mention the coefficients of the lemniscate function studied by Hurwitz [4].

DUKE UNIVERSITY.

REFERENCES.

[1] L. Carlitz, "Congruences for the coefficients of the Jacobi elliptic functions," *Duke Mathematical Journal*, vol. 16 (1949), pp. 297-302.

[2] ———, "Kummer congruences and the Schur derivative," *American Journal of Mathematics*, vol. 75 (1953), pp. 699-706.

[3] ———, "Some theorems on Kummer's congruences," *Duke Mathematical Journal*, vol. 20 (1953), pp. 423-431.

[4] A. Hurwitz, "Über die Entwickelungs coeffizienten der lemniscatischen Functionen," *Mathematische Annalen*, vol. 51 (1899), pp. 196-226 (= *Mathematische Werke*, Basel, 1933, vol. 2, pp. 342-373).

[5] H. S. Vandiver, "Note on a certain ring congruence," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 418-423.

ON THE FUNCTIONAL EQUATION
$dy/dx = f(x, y(x), y(x+h)), h > 0$.*

By SHAFIK DOSS and SAAD K. NASR.

1. The object of the present note is to show that if certain conditions bearing on the function $f(x, y, z)$ of three variables are satisfied, then there is just one bounded solution of the given functional equation, defined for $x \geqq x_0$ and satisfying the initial condition $y(x_0) = y_0$.

To prove the statement we shall apply Picard's method of successive approximations, making the following assumptions which will insure the convergence of the process:

(i) $\quad | f(x, Y, Z) - f(x, y, z) | \leqq K_1(x) | Y - y | + K_2(x) | Z - z |$,

(ii) $\quad \int_{x_0}^{\infty} (K_1(x) + K_2(x)) dx = B < 1$,

(iii) there exist a y and a z for which $\int_{x_0}^{\infty} | f(t, y, z) | dt < \infty$.

Condition (i) is a Lipschitz hypothesis of a very strong type. If the integral in (ii) is convergent, then condition (ii) can always be satisfied by choosing x_0 large enough. The solution will be defined for $x \geqq x_0$.

With reference to a constant y_0, consider the sequence of functions defined by

$$y_1(x) = y_0 + \int_{x_0}^{x} f(t, y_0, y_0) dt,$$

$$y_n(x) = y_0 + \int_{x_0}^{x} f(t, y_{n-1}(t), y_{n-1}(t+h)) dt \qquad n \geqq 2.$$

We have $| y_1(x) - y_0 | \leqq \int_{x_0}^{x} | f(t, y_0, y_0) | dt \leqq A$, where A is finite by (i), (ii) and (iii). If we assume

$$| y_{p-1}(x) - y_{p-2}(x) | \leqq AB^{p-2}, \qquad 2 \leqq p \leqq n, \qquad y_0(x) = y_0,$$

* Received December 15, 1952; revised April 30, 1953.

then

$$| y_n(x) - y_{n-1}(x) | \leqq \int_{x_0}^{x} \{K_1(t) | y_{n-1}(t) - y_{n-2}(t) | + K_2(t) | y_{n-1}(t+h) - y_{n-2}(t+h) | \} dt \leqq AB^{n-1}.$$

The series $\Sigma(y_n(x) - y_{n-1}(x))$ is therefore absolutely and uniformly convergent and the function $y(x) = \lim_{n\to\infty} y_n(x)$ exists and satisfies

$$(1) \qquad y(x) = y_0 + \int_{x_0}^{x} f(t, y(t), y(t+h))\, dt,$$

which is equivalent to the given functional equation with the initial condition $y(x_0) = y_0$. We observe that $| y(x) | \leqq | y_0 | + A/(1-B)$.

It is now possible to prove that $y(x)$ is the only *bounded* solution $Y(x)$ which satisfies (1). In fact, if $Y(x)$ is such a solution of (1), then

$$Y(x) = y_0 + \int_{x_0}^{x} f(t, Y(t), Y(t+h))\, dt,$$

and we can easily deduce, with the previous notations and in virtue of conditions (i) and (ii), that

$$| Y(x) - y_n(x) | \leqq MB^n, n \geqq 1, \text{ where } M = \max_t \{ | Y(t) - y_0 | \}.$$

We therefore get $Y(x) = \lim_{n\to\infty} y_n(x) = y(x)$ and our statement is proved.

If we now assume that condition (i) is satisfied for Y, y, Z, z belonging to the interval $(y_0 - b, y_0 + b)$, then, provided $A/(1-B) < b$ holds (and this is always possible if x_0 is taken large enough), we get by induction

$$| y_n(x) - y_{n-1}(x) | \leqq AB^n \text{ and } | y_n(x) - y_0 | < A/(1-B) < b.$$

Therefore, the solution $y(x) = \lim_{n\to\infty} y_n(x)$ exists and satisfies

$$(2) \qquad | y(x) - y_0 | < b,$$

and just as before we may show that this is the only solution satisfying (2).

Particular case. For the equation

$$(3) \qquad dy/dx = A(x)y(x) + B(x)y(x+h), h > 0,$$

the previous considerations will apply if we assume that

$$(4) \qquad \int^{\infty} | A(x) | \, dx \text{ and } \int^{\infty} | B(x) | \, dx \text{ are convergent.}$$

Hence (3) has, under the assumptions (4), just one bounded solution defined for $x \geqq x_0$ and satisfying the initial condition $y(x_0) = y_0$.

2. We owe to a referee the following corollaries and comments.

(a) If $y(x)$ is a bounded solution, then $\lim y(x) = \lambda$ exists as $x \to \infty$.

(b) The limit λ can have any real value, and the correspondence $y \leftrightarrow \lambda$ is one-to-one.

(c) If $x_0 = -\infty$, there still is a one-parameter family of bounded solutions; for every bounded solution both $y(-\infty)$ and $y(+\infty)$ exist; each one of these can be chosen arbitrarily and determines the solution uniquely; the correspondence $y(-\infty) \leftrightarrow y(+\infty)$ is one-to-one again.

(d) If the integral in (ii) is not convergent, the theorem is no longer true.

(e) The case of the linear homogeneous equation (3) is a consequence of the following stronger theorem (which seems to be classical):

If $\int_0^\infty | K(x,t) | \, dt \leqq C < 1$, then $y(x) = y_0 + \int_0^\infty K(x,t) y(t) \, dt$ has just one bounded solution.

To show that, for a bounded solution $y(x)$, the limit $y(\infty)$, exists, consider $y(x)$ as given by (1) and let $| y(x) | < M$. We have, in virtue of (i), (ii) and (iii),

$$\int_{x_0}^{x} | f(t, y(t), y(t+h)) | \, dt \leqq \int_{x_0}^{x} | f(t, y, z) | \, dt$$
$$+ \, | M - y | \int_{x_0}^{x} K_1(t) \, dt + | M - z | \int_{x_0}^{x} K_2(t) \, dt < A,$$

where A is finite and independent of x. This shows that

$$\int_{x_0}^{\infty} f(t, y(t), y(t+h)) \, dt$$

is convergent and $\lambda = \lim_{x \to \infty} y(x)$ exists and is finite.

To show (b), consider a bounded solution $Y(x)$ of

$$Y(x) = Y_0 + \int_{x_0}^{x} f(t, Y(t), Y(t+h)) \, dt.$$

It is easy to see that

$$(5) \qquad | Y_0 - y_0 | / (1 + B) \leqq | Y(x) - y(x) | \leqq | Y_0 - y_0 | / (1 - B).$$

If we observe that, in virtue of our uniqueness theorem, λ is a non-decreasing function of y_0, we see that (5) implies property (b).

The previous considerations apply if $x_0 = -\infty$ and property (c) is therefore true.

To prove property (d), consider

$$(6) \qquad y'(x) = (y(x) - y(x+h))/(e^{-h} - 1).$$

A solution of (6) is $y(x) = e^{-x}$. For $x \geqq x_0$, the functions $y(x) = ae^{-x} + b$ form a family of bounded solutions depending on two arbitrary constants. There is thus an infinity of bounded solutions satisfying the initial condition $y(x_0) = y_0$.

To show (e), we only have to define $K(x, t)$, for $x \geqq 0$, $t \geqq 0$, as follows:

$$K(x,t) = A(t) \text{ for } x \geqq t,\ t < h;\ K(x,t) = A(t) + B(t-h) \text{ for } x \geqq t \geqq h;$$
$$K(x,t) = B(t-h) \text{ for } 0 \leqq t-h < x < t;\ K(x,t) = 0 \text{ for } x < t < h;$$
$$K(x,t) = 0 \text{ for } x \leqq t-h.$$

Then any bounded solution of (3) satisfying the initial condition $y(0) = y_0$ will be a bounded solution of

$$(7) \qquad y(x) = y_0 + \int_0^\infty K(x,t)y(t)\,dt.$$

That (7) has, under the assumption $\int_0^\infty |K(x,t)|\,dt \leqq C < 1$, only one bounded solution, can be seen, just as in the proof of our theorem, by the consideration of the sequence of functions

$$y_1(x) = y_0 + \int_0^\infty K(x,t)y_0\,dt, \quad y_n(x) = y_0 + \int_0^\infty K(x,t)y_{n-1}(t)\,dt, \quad n \geqq 2.$$

Property (e) is therefore proved.

FACULTY OF SCIENCE,
ALEXANDRIA UNIVERSITY, EGYPT.

ON NON-OSCILLATORY LINEAR DIFFERENTIAL EQUATIONS.*

By Philip Hartman and Aurel Wintner.

There are various senses in which the linear differential equation

$$(1) \qquad x'' + f(t)x = 0,$$

where $f(t)$ is a given continuous function defined for large positive t, can be required to have the same asymptotic behavior ($t \to \infty$) as the case $f(t) \equiv 0$ of (1). First, since $x'' + 0x = 0$ has the two solutions $x = x(t) \equiv 1$, $x = y(t) \equiv t$, the simplest condition under which $f(t)$ may be considered to be *small* ($t \to \infty$) in (1) appears to be the following requirement: (1) has some pair of solutions $x = x(t)$, $y = y(t)$ satisfying

$$(2_1) \qquad x(t) \to 1; \qquad\qquad (2_2) \qquad y(t) \sim t,$$

as $t \to \infty$. A stronger requirement is the existence of a pair of solutions for which these asymptotic relations hold and for which the derivatives satisfy

$$(3_1) \qquad x'(t) \to 0; \qquad\qquad (3_2) \qquad y'(t) = o(t).$$

A still stronger requirement is the pair of relations

$$(4_1) \qquad x'(t) = o(t^{-1}); \qquad\qquad (4_2) \qquad y'(t) \to 1,$$

the second of which, being the result of a formal differentiation of (2_2), is a natural desideratum, whereas (4_1), being a substantial refinement of the formal derivative (3_1) of (2_1), appears to be quite artificial. But this appearance is misleading, since it turns out that

(5) the existence of a solution $y = y(t)$ satisfying (4_2) is equivalent to the existence of a solution $x = x(t)$ satisfying (4_1). In addition,

(6) the existence of a solution $y = y(t)$ satisfying both (3_2) and (2_2) is equivalent to the existence of a solution satisfying both (3_1) and (2_1). Finally,

(7) the existence of a solution $x = y(t)$ satisfying (2_2) is equivalent to the existence of a solution $x = x(t)$ satisfying (2_1). Since it is clear (by integration) that

(8) condition (4_2) [but not (3_2)] implies (2_2),

* Received March 5, 1953.

it follows that *the assumptions* (3_2), (4_1) *are equivalent to the formal differentiability of the respective asymptotic relations* (2_1), (2_2). Note that, in view of (7) and of the superposition principle, (1) can never have two solutions satisfying (2_1), simply because exactly one of the two solutions (2_1), (2_2) is bounded as $t \to \infty$ (so that any two such solutions are linearly independent).

In order to prove (5) and (6), grant first (7) and note that, since (2_1) and (2_2) are two linearly independent solutions of (1), their Wronskian is a non-vanishing constant,

$$x(t)y'(t) - y(t)x'(t) = \text{const.} \neq 0. \tag{9}$$

Both assertions of (5) and both assertions of (6) follow immediately if each of the four pairs of asymptotic relations assumed in (5) and (6) is substituted into (9).

In order to prove (7), note that, if $x = u(t)$ is any non-vanishing solution of (1), then direct differentiations show that $x = v(t)$ and $x = w(t)$, where

$$(10_1)\quad v(t) = u(t)\int^{t} (u(s))^{-2}ds; \qquad (10_2)\quad w(t) = u(t)\int_{t}^{\infty} (u(s))^{-2}ds,$$

are also solutions of (1), with the understanding that $w(t)$ exists only if the integral occurring in (10_2) is convergent. The assertion of (7) follows if (10_1) is applied to $u(t) = x(t)$ or (10_2) to $u(t) = y(t)$ according as (2_1) or (2_2) is assumed.

This proves that the case $k = 1$ of any of the three properties (2_k), $(2_k) + (3_k)$, $(2_k) + (4_k)$ of (1) or $f(t)$ implies and is implied by the corresponding case $k = 2$. Hence it is sufficient to consider each of these three properties for the case $k = 1$ alone. The first of these three properties consists of the existence of a solution curve $x = x(t)$ which, as $t \to \infty$, tends to a line λ parallel to, but distinct from, the t-axis of the (t, x)-plane; cf. (2_1). The second and third properties require, besides the first property, that, as $t \to \infty$, the line λ should become a limit tangent and an asymptote, respectively; cf. (3_1) and (4_1). In fact, the definition of a curve $x = x(t)$ having an asymptote is the existence of a finite limit

$$\lim_{t\to\infty} (x(t) - tx'(t)), \tag{11}$$

which, if (2_1) or the existence of a finite $x(\infty) = \lim x(t)$ is assumed, is equivalent to (4_1).

It will be convenient to introduce the following terminology: (1) or $f(t)$ will be said to have the property (2), (3) or (4) if (1) has solutions $x=x(t)$, $x=y(t)$ satisfying (2), (2) and (3), or (2) and (4), respectively.

In what follows, sufficient conditions, and also necessary conditions, will be proved for an f having property (4), and also for an f having the weaker property (3). On the other hand, there will result no conditions which are sufficient for property (2) without being sufficient for (3) too.

(i) *In order that* (1) *has property* (3), *it is sufficient that the improper integral*

$$\int^{\infty} f(t)\,dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty}\int^{T}\right) \tag{12}$$

(*possibly just conditionally*) *and that the corresponding indefinite integral*

$$F(t) = \int_t^{\infty} f(s)\,ds \qquad \left(\int_t^{\infty} = \lim_{T\to\infty}\int_t^{T}\right) \tag{13}$$

satisfies the following conditions:

$$\int^{\infty} F(t)\,dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty}\int^{T}\right) \tag{14}$$

(*possibly just conditionally*) *and*

$$\int^{\infty} tF^2(t)\,dt < \infty. \tag{15}$$

The following partial converse of (i) will be clear from the proof of (i):

(i bis) *If conditions* (12)-(14) *are satisfied, then* (15) *is not only sufficient but necessary as well in order that* (1) *has property* (3).

Remark. It follows from (i bis) that (14) and (15) in (i) cannot be replaced by

$$\int^{\infty} |F(t)|\,dt < \infty. \tag{16}$$

The following counterpart of (i) will also be proved:

(ii) *In order that* (1) *has property* (4), *it is sufficient that the conditions* (12)-(15) *of* (i) *hold and that*

$$F(t) = o(t^{-1}) \qquad (t\to\infty). \tag{17}$$

What now corresponds to (i bis) is the following fact:

(iii) *In order that* (1) *possesses property* (4), *it is necessary that* (12) *holds and that the function* (13) *satisfies* (17).

REMARK. If (12) is satisfied by (1), that is, if the function $F(t)$ exists (which, according to (iii), is necessary for property (4)), and *if* $F(t)$ *does not change sign from a certain* $t = t_0$ *onward* (which is the case if, but not only if, $f(t)$ does not change sign from a certain $t = t^0$ onward), *then not only* (17) *but also* (16) *is necessary in order that* (1) *has property* (4). This can be seen as follows: According to (1) and (13),

$$(x - tx')' \equiv - tx'' = tfx = - txF' \equiv (- txF)' + (x + tx')F.$$

If a quadrature is applied to this identity, then, since (4) and (17) imply that $- txF = o(1)$, it follows that the limit (11) exists if and only if the integral

$$\int^{\infty} (x + tx')Fdt \text{ converges} \qquad (\int^{\infty} = \lim_{T\to\infty} \int^{T}).$$

In view of (4), this proves the last italicized assertion.

A set of sufficient conditions, different from those supplied by (i)-(ii), for (1) to possess the property (4) is contained in the following theorem:

(iv) *In order that* (1) *has property* (4), *it is sufficient that*

$$\int^{\infty} tf(t)\,dt \text{ converges} \qquad (\int^{\infty} = \lim_{T\to\infty} \int^{T}) \tag{18}$$

and

$$\int^{\infty} t^{2p-1} \,|\, f(t)\,|^{p} dt < \infty, \tag{19}$$

where p *is some index on the range*

$$1 \leqq p \leqq 2. \tag{20}$$

The assumptions (12)-(15) and (17) of (i) and (ii) do not require that

$$\int^{\infty} |\, f(t)\,|\, dt < \infty, \tag{21}$$

nor that (16) holds, and still less that

$$\int^{\infty} \max_{t \leqq s < \infty} |\, F(s)\,|\, dt < \infty \tag{22}$$

or (what is still more stringent) that

$$(23) \qquad \int^{\infty} t\,|f(t)|\,dt < \infty.$$

On the other hand, (23) or (12) and (22) imply the conditions (12)-(15) and (17) of (i) and (ii). In fact, both (22) and (23) imply (22) (hence (16)) and (17), while (16) and (17) imply (15). Thus (i) improves a classical result (cf. [5], p. 486, footnotes 57 and 58, and [8], p. 854, footnote), in which (23) is assumed, and a refinement of this result ([7], p. 595), in which (12) and (22) are assumed, and in both of which results it is asserted that (1) has the property (3). Actually, the literature consulted does not seem to point out, even under the assumption (23), that the conclusion of (i) can be strengthened to that of (ii). It is, of course, easy to see that (1), (2_1) and (23) imply (4_1).

In the course of the proof of (i), where it will be assumed that (12) holds, it will be convenient to introduce the functions

$$(24) \qquad G(t) = H^2(t)\int^t ds/H^2(s), \text{ where } H(t) = \exp\int^t F(s)\,ds.$$

Since (14) implies that

$$(25) \qquad H(t) \to \text{const.} \ (>0) \text{ and } G(t) \sim t \qquad (t\to\infty),$$

it is clear that (i), (i bis) and (ii) are contained in the following theorem:

(v) *If* (12) *is assumed, and if* $G(t)$ *and* $H(t)$ *are defined by* (24), *then* (1) *possesses a solution* $x = x(t)$ *and/or a solution* $x = y(t)$ *satisfying*

$$(26_1) \qquad x \sim H(t); \qquad\qquad (26_2) \qquad y \sim G(t)/H(t)$$

$(t\to\infty)$ *if and only if*

$$(27) \qquad \int^{\infty} dt/H^2(t) = \infty$$

and

$$(28) \qquad \int^{\infty} G(t)F^2(t)\,dt < \infty;$$

in which case the relations

$$(29_1) \quad x'/x = F + o(G^{-1}); \qquad\qquad (29_2) \quad y'/y = F + (1+o(1))G^{-1}$$

also hold.

REMARK. It follows from a criterion of [9], pp. 376-377, that *assumption* (27) *of* (v) *is satisfied whenever* (1) *is non-oscillatory.*

(iv) is contained in Theorems (vi) and (vii) below.

(vi) *In order that* (1) *posssesses a solution* $x = x(t)$ *and/or a solution* $x = y(t)$ *satisfying*

$$(30_1) \qquad x'/x = o(t^{-1}); \qquad\qquad (30_2) \qquad y'/y \sim t^{-1}$$

$(t \to \infty)$, *it is necessary and sufficient that*

$$(31) \qquad \underset{0<s<\infty}{\text{l. u. b.}} \left| \int_t^{t+s} rf(r)\,dr \right| / (1 + \log(1 + s/t)) \to 0 \quad \text{as} \quad t \to \infty.$$

It is easily verified that (31) is satisfied if either (18) holds or (19) holds for some $p \geqq 1$. It is understood that (vi) is meant to imply that (31) is sufficient for (1) to be non-oscillatory (that is, in order that no solution $x = x(t) \not\equiv 0$ of (1) can vanish for large t-values).

(vii) *If* (19) *and* (20) *hold, then* (1) *possesses a pair of solutions* $x = x(t)$, $x = y(t)$ *satisfying*

$$(32_1) \qquad x \sim \exp \int^t sf(s)\,ds; \qquad\qquad (32_2) \qquad y \sim t \exp - \int^t sf(s)\,ds$$

and (30_1), (30_2), *respectively.*

This furnishes a criterion for (1) to have the property (2), as well as for the following situation: (1) has solutions $x = x(t)$, $y = y(t)$ satisfying

$$(33_1) \qquad 0 \not\equiv x = o(1); \qquad\qquad (33_2) \qquad yt^{-1} \to \infty.$$

COROLLARY. *If* (19) *and* (20) *hold and if*

$$(34) \qquad \int^t sf(s)\,ds \to -\infty, \qquad (t \to \infty),$$

then (1) *has solutions* $x = x(t) \not\equiv 0$ *and* $x = y(t)$ *which satisfy* (33_1), (30_1) *and* (33_2), (30_2), *respectively.*

The above theorems will be proved in the following order: (v), (vi), (vii), (iii).

Proof of (v). With a fixed choice of the lower limit of integration, put

$$(35) \qquad z = z(t) = \exp \int^t F(s)\,ds.$$

Then $z(t)$ satisfies the differential equation

$$z'' + (f - F^2)z = 0. \tag{36}$$

Thus, if $x = x(t)$ is a solution of (1) and if $u = u(t)$ is defined by

$$x = zu, \tag{37}$$

then u satisfies the differential equation

$$(z^2u')' + F^2z^2u = 0; \tag{38}$$

conversely, if $u = u(t)$ is a solution of (38), then (37) is a solution of (1). Thus, (1) has a solution $x = x(t)$ satisfying (26_1) if and only if (38) has a solution satisfying

$$u \to 1 \qquad (t \to \infty). \tag{39}$$

Introduce the new independent variable

$$\tau = \int^t z^{-2}(s)\,ds = \int^t (\exp -2\int^s F(r)\,dr)\,ds \tag{40}$$

and let $\dot{u}$ denote $du/d\tau$. Then (27) means that $\tau \to \infty$ as $t \to \infty$. The differential equation (38) becomes

$$\ddot{u} + F^2z^4u = 0 \qquad (t = t(\tau)). \tag{41}$$

Since $F^2z^4 \geqq 0$, a necessary and sufficient condition in order that (41) has a solution $u = u(\tau)$ satisfying (39) as $\tau \to \infty$ (or $t \to \infty$) is that

$$\int^\infty \tau F^2z^4\,d\tau < \infty; \tag{42}$$

cf. [7], Appendix, pp. 601-602. (The condition (42) on (41) is the analogue of condition (23) on (1).) In view of (24), (35) and (40), the inequality (42) is precisely (28).

When (28), that is, (42), holds, then (41) has a pair of solutions $u = u(\tau)$, $u = v(\tau)$ satisfying

$$(43_1) \quad u \to 1, \quad \dot{u} = o(\tau^{-1}); \qquad (43_2) \quad v \sim \tau, \quad \dot{v} \to 1,$$

as $\tau \to \infty$. Furthermore, the existence of one of these solutions is equivalent to the existence of the other; cf. (5).

Consider the solutions $x = zu$ and $y = zv$ of (1). Clearly, (26_1) and (26_2) hold. Since $x'/x = z'/z + \tau'\dot{u}/u$ and $y'/y = z'/z + \tau'\dot{v}/v$, it follows from (40) and (43), that (29_1) and (29_2) hold. Thus, if (12) is assumed,

(27) and (28) are sufficient for the existence of a solution $x=x(t)$ of (1) satisfying (26_1), (29_1) and/or for the existence of a solution $x=y(t)$ satisfying (26_2), (29_2).

There remains to be proved the necessity of (27) and (28). If (1) has a solution satisfying (26_1) or (26_2), then (1) is non-oscillatory. This implies (27) (cf. [9], pp. 376-377; for a refinement of this assertion, cf. [3]). Finally, in the above proof of the sufficiency of (28), it was pointed out that (28) is necessary and sufficient when (12) and (27) hold. This completes the proof of (v).

Proof of (vi) *and* (vii). Consider the variation of constants and the change of independent variables

$$x=e^{\tau}u \tag{44}$$

and

$$\tau=\tfrac{1}{2}\log t \qquad (t=e^{2\tau}), \tag{45}$$

respectively; so that

$$\tau'=\tfrac{1}{2}t^{-1}=\tfrac{1}{2}e^{-2\tau}; \qquad \dot{t}=2t=2e^{2\tau}, \tag{46}$$

if $\dot{u}=du/d\tau$. By virtue of (44) and (45), the differential equation (1) is equivalent to

$$\ddot{u}-(1+\phi(\tau))u=0, \tag{47}$$

where

$$\phi(\tau)=-4t^2f(t), \qquad t=e^{2\tau}. \tag{48}$$

A necessary and sufficient condition that (1) be non-oscillatory and possess a solution $x=x(t)$ $[x=y(t)]$ satisfying (30_1) $[(30_2)]$ is that (47) be non-oscillatory and have a solution $u=u(\tau)$ $[u=v(\tau)]$ satisfying

$$(49_1)\qquad \dot{u}\sim -u; \qquad\qquad (49_2)\qquad \dot{v}\sim v.$$

This is clear if $x=e^{\tau}u$ and $y=e^{\tau}v$, since $x'/x=\tau'(1+\dot{u}/u)$ and $y'/y=\tau'(1+\dot{v}/v)$, where $\tau'=\tfrac{1}{2}t^{-1}$. On the other hand, a necessary and sufficient condition for (47) to be non-oscillatory and to have a solution $u=u(\tau)$ and/or a solution $u=v(\tau)$ satisfying (49_1) or (49_2), respectively, is that

$$\operatorname*{l.\,u.\,b.}_{0<\sigma<\infty} \left|\int_{\tau}^{\tau+\sigma}\phi(\rho)\,d\rho\right| / \left(1+\int_{\tau}^{\tau+\sigma}d\rho\right)\to 0, \tag{50}$$

as $\tau\to\infty$; cf. [2], Theorem (IV), p. 570. Since (50) is equivalent to (31) by virtue of (48), assertion (vi) follows.

In order to prove (vii), note that (19) is equivalent to

$$(51) \qquad \int^{\infty} |\phi(\tau)|^p \, d\tau < \infty,$$

by (48). By [2], Theorem (VII), p. 575, condition (51) assures the existence of solutions $u = u(\tau)$, $v = v(\tau)$ of (47) satisfying, as $\tau \to \infty$,

$$(52_1) \qquad u(\tau) \sim \exp(-\tau - \tfrac{1}{2} \int^{\tau} \phi(\sigma)\, d\sigma);$$

$$(52_2) \qquad v(\tau) \sim \exp(\tau + \tfrac{1}{2} \int^{\tau} \phi(\sigma)\, d\sigma)$$

and (49_1), (49_2). The corresponding solutions $x = e^{\tau}u$ and $y = e^{\tau}v$ of (1) satisfy (30) and (32). This proves (vii).

Reversing the procedures of the proofs of (vi) and (vii), it is possible to obtain conclusions concerning the asymptotic behavior of solutions of (47) from those concerning solutions of (1). For example, (i) and (v) supply the following criterion:

Let $\phi(\tau)$ be continuous for large τ and have the properties that

$$(53) \qquad \int^{\infty} \phi(\tau) e^{-2\tau} d\tau \text{ converges} \qquad (\int^{\infty} = \lim_{T\to\infty} \int^{T})$$

(possibly just conditionally) and that the corresponding indefinite integral

$$(54) \qquad \Phi(\tau) = \int_{\tau}^{\infty} \phi(\sigma) e^{-2\sigma} d\sigma \qquad (\int^{\infty} = \lim_{T\to\infty} \int^{T})$$

satisfies the following conditions:

$$(55) \qquad \int^{\infty} \Phi(\tau)\, d\tau \text{ converges} \qquad (\int^{\infty} = \lim_{T\to\infty} \int^{T})$$

(possibly just conditionally) and

$$(56) \qquad \int^{\infty} e^{4\tau}\Phi^2(\tau)\, d\tau < \infty.$$

Then (47) possesses a pair of solutions $u = u(\tau)$, $v = v(\tau)$ satisfying

$(57_1) \qquad u \sim e^{-\tau}; \qquad\qquad (57_2) \qquad v \sim e^{\tau}$

and

$$(58_1) \qquad \dot{u}/u = -1 + O(|\Phi(\tau)|e^{2\tau}) + o(1);$$

$$(58_2) \qquad \dot{v}/v = 1 + O(|\Phi(\tau)e^{2\tau}) + o(1).$$

Proof. By (45) and (48), the relations (12), (14), (15) are respectively identical with (53), (55), (56); while, by (44), the existence of solutions $x = x(t)$, $x = y(t)$ of (1) satisfying (3) is equivalent to the existence of solutions of (47) satisfying (57). Finally, (58) follows from (29) by virtue of (44), (45).

Similarly, (v) itself leads to theorems concerning asymptotic integrations of (47) in which (57), the analogue of (2), is replaced by

$$(59_1) \qquad u \sim \exp(-\tau + \int^{\tau} \Phi(\sigma)d\sigma);$$

$$(59_2) \qquad v \sim \exp(-\tau + \int^{\tau} \Phi(\sigma)d\sigma) \cdot \int^{\tau} (\exp -2\int^{\sigma} \Phi(\rho)d\rho)d\sigma.$$

Proof of (iii). If a solution $x = x(t)$ of (i) does not vanish (on a t-interval), then $l = x'/x$ exists and satisfies the Riccati differential equation $l' + l^2 + f(t) = 0$ (on that t-interval). If $x = x(t)$ satisfies (2_1), then

$$(60) \qquad l(t) + \int^{t} l^2(s)ds + \int^{t} f(s)ds = l(t_0),$$

if the lower limit of integration $t = t_0$ is sufficiently large and fixed. If, in addition, $x = x(t)$ satisfies (4_1), then $l(t) \to 0$, and the first integral in (60) tends to a finite limit, as $t \to \infty$. Hence (12) holds. Consequently, (60) can be written as

$$(61) \qquad l(t) = \int_t^{\infty} l^2(s)ds + F(t).$$

In view of (4_1), this implies that $o(t^{-1}) = o(t^{-1}) + F(t)$, which proves (13).

This completes the proofs of all assertions made above.

THE PROPERTY (4^*). In what follows, differential equations (1) will be considered for which the existence of a solution $x = y(t)$ satisfying (4_2) can be improved to

$$(62) \qquad y(t) = t + o(1), \qquad y'(t) = 1 + o(t^{-1}).$$

It turns out that this is the case if and only if *every solution of* (1) *has an asymptote,* that is, if and only if (1) has two linearly independent solutions $x = x_1(t)$, $x = x_2(t)$ each of which leads to a finite limit (11). Then (1) will be said to have *property* (4*). It will be seen below that property (4) is necessary, though not sufficient, for property (4*).

The following remarks (α), (β) will be needed:

(α) If $x = x(t)$ is any fixed solution of (1), it has an asymptote if and only if the improper integral

$$(63) \qquad \int^{\infty} tf(t)x(t)\,dt \text{ converges} \qquad \left(\int^{\infty} = \lim_{T\to\infty} \int^{T}\right).$$

(β) A function $x = x(t)$, possessing a continuous first derivative $x'(t)$ for large positive t, has an asymptote if and only if there exists a pair of constants a, b satisfying both conditions

$$(64_{\mathrm{I}}) \quad x(t) = at + b + o(1); \qquad (64_{\mathrm{II}}) \quad x'(t) = a + o(t^{-1}) \qquad (t \to \infty).$$

Proof of (α). Since (1) means that $(x - tx')' = -tx''$ is identical with tfx, the existence of the limit (11) is equivalent to condition (63).

Proof of (β) (cf. [1], pp. 144-145). The existence of the limit (11) means that $x = x(t)$ satisfies a linear differential equation of the form $x = tx' + \text{const.} + h(t)$, where $h(t)$ is a (continuous) function satisfying $h(t) \to 0$ as $t \to \infty$. If this linear differential equation of first order is integrated by a quadrature, (64_{I}) follows. But direct substitutions show that if $x(t)$ is of the form (64_{I}), then (64_{II}) is equivalent to the existence of the limit (11).

It is easy to conclude from (β) that property (4*) implies property (4). In fact, if there belongs to every solution $x(t)$ of (1) a pair of constants a, b satisfying (64_{I}) and (64_{II}), then $a \neq 0$ for some solution. For, if it is assumed that (64_{I}) and (64_{II}) hold for two linearly independent solutions $x = x_1(t)$, $x = y_2(t)$ with $(a, b) = (0, b_1)$, $(a, b) = (0, b_2)$, respectively, then the Wronskian $x_1x_2' - x_2x_1'$ is $O(1)o(1) + O(1)o(1) = o(1)$, which is impossible, since the Wronskian of two linearly independent solutions of (1) is a non-vanishing constant. Accordingly, property (4*) implies the existence of a solution $x = y(t)$ for which (64_{I}) and (64_{II}) hold with $a = 1$. But this implies (4_2), and therefore the existence of a solution $x = x(t)$ satisfying (2_1) and (4_1).

The equivalence of property (4*) and of the existence of a solution

$y=y(t)$ satisfying (62) now follows by taking a linear combination of solutions satisfying (2_1)-(4_1) and the case $a=1$ of (64_{I})-(64_{II}).

In particular, (1) must possess a solution $x(t)\sim t$. It follows that *if $f(t)$ does not change sign from a certain $t=t^0$ onward* (that is, if

$$f(t)=\pm\,|\,f(t)\,|, \tag{65}$$

where the choice of the sign is independent of $t\geqq t^0$), *then the condition*

$$\int^{\infty} t^2\,|\,f(t)\,|\,dt<\infty \tag{66}$$

is necessary and sufficient in order that (1) *has property* (4*). In fact, if (1) has property (4*), then (63) must be satisfied by every solution $x(t)$, and therefore by some solution $x(t)\sim t$, and so the necessity of (66) follows from the assumption (65). On the other hand, the sufficiency of (66) for the (4*)-character of (1) is true without the assumption (65) also. This follows as a particular case ($g\equiv 0$) of the following criterion (γ):

(γ) If $f(t)$ and $g(t)$ are continuous functions satisfying (66) and

$$\int^{\infty} t\,|\,g(t)\,|\,dt<\infty, \tag{67}$$

then every solution $x=x(t)$ of

$$x''+g(t)x'+f(t)x=0 \tag{68}$$

has an asymptote.

This criterion (γ) is known; cf. [4], where, however, a lengthy proof is given, since it is not observed that (γ) is an easy consequence of the following lemma (γ bis) which, in another form, is of a much older date (cf. [5], p. 486, footnotes 57 and 58, and [8], p. 854, footnote; for a simple proof, cf. [6], pp. 262-264).

(γ bis) Let the coefficients $a_{ik}(t)$ of a linear differential system

$$u'=a_{11}(t)u+a_{12}(t)v, \qquad v'=a_{21}(t)u+a_{22}(t)v \tag{69}$$

be continuous functions satisfying

$$\int^{\infty} |\,a_{ik}(t)\,|\,dt<\infty \qquad (i,k=1,2). \tag{70}$$

Then there belongs to every pair of constants α, β a unique solution of (69)

satisfying $(u(t), v(t)) \rightarrow (\alpha, \beta)$ as $t \rightarrow \infty$ (which implies that the limits $\alpha = u(\infty)$, $\beta = v(\infty)$ exist for every solution).

In order to deduce (γ) from $(\gamma \text{ bis})$, put, with reference to any solution $x = x(t)$ of (68),

$$(71) \qquad u = x - tx', \qquad v = x'.$$

Then $u' = -tx'' = t(gx' + fx) = tgv + tf(tv + u)$, $v' = x'' = -u'/t$. This representation of u', v' can be written in the form (69), with

$$(72) \qquad a_{11} = tf, \quad a_{12} = tg + t^2 f, \quad a_{21} = -f, \quad a_{22} = -g - tf.$$

Since (70) can be reduced to the pair of conditions (66)-(67) in the case (72), it is seen from the definition (71) of u that (γ) is a corollary of $(\gamma \text{ bis})$.

Remark. Suppose that (1) is oscillatory, that is, that some (hence, according to Sturm, every) solution $x(t) \not\equiv 0$ of (1) has zeros $t = t_n$ which cluster at $t = \infty$. Then (1) *can* have a solution $x(t) \not\equiv 0$ possessing an asymptote (cf. below), but it cannot have two, linearly independent, such solutions. For, if it did, (1) would possess property (4*), hence property (4), and therefore a solution satisfying (2_1), which is impossible when (1) is oscillatory. That *one* solution $(\not\equiv 0)$ possessing an asymptote *is* compatible with the oscillatory character of (1), is shown by the following example:

Draw in the (t, x)-plane a graph $x = x(t)$ in such a way that the graph of $|x(t)|$ consists of a sequence $(t_1, t_2), (t_2, t_3), \cdots$ of strictly convex arches, and that $\operatorname{sgn} x(t) = (-1)^n$ if $t_n < t < t_{n+1}$, while $x(t_n) = 0$ and $x'(t_n) \neq 0$, where $n = 1, 2, \cdots$ and $t_n \rightarrow \infty$. Then, if the function $x(t)$ possesses a continuous third derivative, a continuous function $f(t)$ is defined by placing $f(t) = -x''(t)/x(t)$ or $f(t) = f(t \pm 0)$ according as $t \neq t_n$ or $t = t_n$, since the existence of $\lim x''(t)/x(t)$, as $t \rightarrow t_n$, follows from l'Hopital's rule. With reference to this $f(t)$, the function $x(t)$ is a solution of (1), hence (1) is oscillatory. But it is clear that the successive waves of the graph of $x(t)$ can be chosen in such a way that $t_n x'(t_n) \rightarrow 0$ and $t_{n+1} - t_n <$ const. (hence $t_{n+1}/t_n <$ Const.) as $n \rightarrow \infty$. Then, for reasons of convexity, $tx'(t) \rightarrow 0$ and $x(t) \rightarrow 0$ as $t \rightarrow \infty$, and so the limit (11) exists.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. H. Hardy, "Generalizations of a limit theorem of Mr. Mercer," *Quarterly Journal of Pure and Applied Mathematics*, vol. 43 (1912), pp. 143-150.

[2] P. Hartman, "Unrestricted solution fields of almost-separable differential equations," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 560-580.

[3] ———, "On linear second order differential equations with small coefficients," *American Journal of Mathematics*, vol. 73 (1951), pp. 955-962.

[4] O. Haupt, "Ueber Lösungen linearer Differentialgleichungen mit Asymptoten," *Mathematische Zeitschrift*, vol. 48 (1942-43), pp. 212-220; cf. pp. 289-292.

[5] E. Hilb, Article II B5 (1914) in the *Encyclopädie der mathematischen Wissenschaften*, vol. II_2.

[6] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

[7] ———, "On almost free linear motions," *ibid.*, vol. 71 (1949), pp. 595-602.

[8] ———, "On linear asymptotic equilibria," *ibid.*, vol. 71 (1949), pp. 853-858.

[9] ———, "On the non-existence of conjugate points," *ibid.*, vol. 73 (1951), pp. 368-380.

LINEAR DIFFERENTIAL AND DIFFERENCE EQUATIONS WITH MONOTONE SOLUTIONS.*

By PHILIP HARTMAN and AUREL WINTNER.

1. A corollary of one of the principal theorems to be proved below is the following:

(i) *In the linear, homogeneous differential equation of n-th order*

$$f_0(t)D^n y + \sum_{k=1}^{n} (-1)^{k+1} f_k(t) D^{n-k} y = 0, \qquad (D = d/dt), \tag{1}$$

let the coefficient functions $f_0, \cdots, f_n$ *be continuous on* $0 < t < \infty$ *and let* f_0 *and* $f_2, \cdots, f_n$ *satisfy*

$$f_0 > 0 \text{ and } f_k \geqq 0 \text{ for } k = 2, 3, \cdots, n, \qquad (0 < t < \infty), \tag{2}$$

while $f_1 \gtreqless 0$. *Then* (1) *has at least one solution* $y = y(t)$ *which is positive and non-decreasing for* $0 < t < \infty$ *and, what is more,*

$$y > 0 \text{ and } (-1)^j D^j y \geqq 0 \text{ for every } j < n \qquad (0 < t < \infty). \tag{3}$$

The same is true if $0 < t < \infty$ *is replaced by* $0 \leqq t < \infty$.

If the second of the inequalities (2) is assumed for $k = 1$ also, then it follows from (1) and (3) that the second of the inequalities (3) remains true for $j = n$.

Assertion (i) is a generalization of a theorem of A. Kneser ([5], pp. 178-192), which assumes that $n = 2$ and that (1) is of the form $f_0 D^2 y - f_2 y = 0$, where $f_1 \equiv 0$; for the general case of $n = 2$, cf. [3]. These proofs for $n = 2$ depend on convexity arguments which are not applicable when $n > 2$. Actually, the proof of (i) below will be much simpler than the proofs of [5], [3] for the particular case $n = 2$.

It is easy to conclude that (i) has the following corollary:

(i bis) *If* $f_0 > 0$ *possesses a completely monotone derivative, and if* $f_1, \cdots, f_n$ *are completely monotone, on* $0 < t < \infty$ *(that is, if*

$$f_0 > 0, \quad (-1)^j D^{j+1} f_0 \geqq 0, \quad (-1)^j D^j f_k \geqq 0 \qquad (k = 1, \cdots, n; j = 0, 1, \cdots), \tag{4}$$

* Received March 31, 1953.

where $0 < t < \infty$), *then* (1) *has at least one solution* $y(t) \not\equiv 0$ *which is completely monotone for* $0 < t < \infty$:

$$(5) \qquad (-1)^j D^j y \geqq 0 \text{ for } j = 0, 1, \cdots .$$

The same is true if $0 < t < \infty$ *is replaced by* $0 \leqq t < \infty$.

In other words, if each of the functions $Df_0, f_1, \cdots, f_n$ is representable as a Laplace transform

$$(6) \qquad \int_0^\infty e^{-ts} d\mu(s), \quad \text{where } d\mu(s) \geqq 0,$$

for $0 < t < \infty$ or $0 \leqq t < \infty$ (Hausdorff-Bernstein), and if $f_0 > 0$, then (1) has at least one solution $y = y(t) \not\equiv 0$ which is representable in the same form on the same t-range. (i bis) was proved in the case $D^2 y - f_2 y = 0$ of (1) in [8] and in the general case of $n = 2$ in [3].

(i) and (i bis) have analogues for difference equations:

(ii) *In the linear, homogeneous difference equation of n-th order*

$$(7) \qquad f_0(m)\Delta^n y(m) + \sum_{k=1}^n (-1)^{k-1} f_k(m)\Delta^{n-k} y(m) = 0, \quad (m = 0, 1, \cdots),$$

let the coefficient sequences $f_0(m), f_1(m), \cdots, f_n(m)$ *satisfy*

$$(8) \qquad f_0(m) > 0 \text{ and } f_k(m) \geqq 0 \text{ for } k = 2, \cdots, n, \qquad (f_1 \gtreqless 0),$$

where $m = 0, 1, \cdots$, *and*

$$(9) \qquad f_0(m) - \sum_{k=1}^n f_k(m) > 0 \qquad (m = 0, 1, \cdots).$$

Then (7) *has at least one solution* $y = y(m)$ *satisfying*

$$(10) \qquad y(m) > 0 \text{ and } (-1)^j \Delta^j y(m) \geqq 0 \text{ for } j = 1, \cdots, n-1 \qquad (m = 0, 1, \cdots).$$

It is understood that $\Delta y(m) = y(m+1) - y(m)$, $\Delta^2 y(m) = \Delta(\Delta y(m)) = y_{m+2} - 2y_{m+1} + y_m, \cdots$. The case $n = 2$ of (ii) was proved in [4].

It is worth making the following remarks concerning (ii):

Remark 1. If it is assumed in (ii) that $f_1(m) \geqq 0$, where $m = 0, 1, \cdots$, then (7) and (9) imply that $(-1)^n \Delta^n y(m) \geqq 0$.

Remark 2. If $n > 1$ and if it is assumed in (ii) that $f_n(m) > 0$ for an infinity of m-values, then $-\Delta y(m) > 0$ for every $m \geqq 0$.

Remark 3. If (9) is weakened to $f_0(m) - \sum_{k=1}^{n} f_k(m) \geqq 0$ and $f_0(m) - f_1(m) > 0$, then (ii) remains valid if the first inequality in (10) is correspondingly weakened to $y(m) \geqq 0$ (so that the solution in question can be of trivial type, that is, $y(m) \equiv 0$ for all sufficiently large m).

The analogue of (i bis) is as follows:

(ii bis) *If* (9) *holds and if each of the* $n+1$ *sequences* $\Delta f_0(m)$, $f_1(m), \cdots, f_n(m)$, *where* $m = 0, 1, \cdots$, *is completely montone (that is, if* (9) *and*

$$(11) \qquad (-1)^j \Delta^{j+1} f_0(m) \geqq 0, \quad (-1)^j \Delta^j f_k(m) \geqq 0 \text{ for } k = 1, 2, \cdots, n \qquad (j, m = 0, 1, \cdots)$$

hold), then (7) *has a solution* $y = y(m) > 0$ *which is completely monotone, that is,*

$$(12) \qquad (-1)^j \Delta^j y(m) \geqq 0 \text{ for } j = 0, 1, \cdots \qquad (m = 0, 1, \cdots).$$

The case $n = 2$ of (ii bis) is known [4].

2. Both (i) and (i bis) will be deduced from corresponding theorems on systems of linear differential equations of first order. In order to simplify the notation, the following abbreviations will be used: If $x = (x_1, \cdots, x_n)$ is a vector, then

$$(13) \qquad x \geqq 0 \text{ means that } x_k \geqq 0 \text{ for } k = 1, \cdots, n.$$

Similarly, if $A = (a_{ik})$ is an n by n matrix, then

$$(14) \qquad A \geqq 0 \text{ means that } a_{ik} \geqq 0 \text{ for } i, k = 1, \cdots, n.$$

The analogous abbreviations $x > 0$ and $A > 0$ will also be used. If $p = (p_1, \cdots, p_n)$ is a vector, the product px will represent the vector

$$(15) \qquad px = (p_1 x_1, p_2 x_2, \cdots, p_n x_n).$$

The main theorem on differential equations to be proved is as follows:

(I) *Let* $A = A(t)$ *be an* n *by* n *matrix of continuous functions satisfying*

$$(16) \qquad A(t) \geqq 0 \qquad (0 < t < \infty).$$

Then the system of differential equations

$$(17) \qquad x' = -A(t)x \qquad ('=D=d/dt)$$

has at least one solution $x = x(t) \not\equiv 0$ *satisfying*

$$x(t) \geqq 0 \text{ and } -x'(t) \geqq 0 \qquad (0 < t < \infty). \tag{18}$$

Needless to say, $x(t)$ is defined and continuously differentiable on the *closed* range $0 \leqq t < \infty$ if $A(t)$ is defined and continuous there.

Remark 1. If $x = x(t) = (x_1(t), \cdots, x_n(t)) \not\equiv 0$ is a solution of (17) satisfying (18), then $x_k(\infty) = \lim x_k(t)$, as $t \to \infty$, exists and is non-negative for $k = 1, 2, \cdots, n$, while $-x'_k(t)$ is non-negative. Thus $-x'_k(t)$ is integrable over $1 \leqq t < \infty$. It follows therefore from (16), (17) and (18) that a necessary condition for $x_k(\infty) > 0$ to hold for some k is that the elements $a_{jk}(t)$, where $j = 1, \cdots, n$, of the k-th column of $A(t)$ be integrable over $1 \leqq t < \infty$. This necessary condition is not sufficient, as is shown by the binary system

$$x'_1 = -x_2, \qquad x'_2 = -x_1/t^2 \qquad (0 < t < \infty).$$

In fact, the general solution $x = (x_1, x_2)$ of this system is

$$x = (c_1 t^\alpha + c_2 t^\beta, \; -\alpha c_1 t^{\alpha-1} - \beta c_2 t^{\beta-1}),$$

where $\lambda = \alpha, \beta = \frac{1}{2}(1 \pm 5^{\frac{1}{2}})$ are the roots of the quadratic equation $\lambda(\lambda - 1) - 1 = 0$. But the system has (up to constant factors) only one solution $x = x(t) = (t^\beta, -\beta t^{\beta-1})$, where $\beta = \frac{1}{2}(1 - 5^{\frac{1}{2}}) < 0$, satisfying (18). For this solution, $x(\infty)$ is $(0, 0)$, although the elements of the first column, $a_{11}(t) \equiv 0$ and $a_{21} = t^{-2}$, are integrable over $1 \leqq t < \infty$.

On the other hand, *when* (16) *holds, a necessary and sufficient condition in order that* (17) *possess a solution satisfying* (18) *and* $x(\infty) > 0$ *is that every element* $a_{jk}(t)$, *where* $j, k = 1, \cdots, n$, *of* $A(t)$ *be integrable over* $1 \leqq t < \infty$. The necessity of this condition follows from the above remarks concerning $x_k(\infty) > 0$ (for some k). The sufficiency of the condition follows from a particular case of a theorem of Dunkel [1]; for a short proof of this particular case, cf. [9], pp. 262-264.

Remark 2. Let $A = A(t)$ in (I) be constant (independent of t) and let $x = x(t) \not\equiv 0$ be a solution of (17) satisfying (18). Every solution of (17), in particular, the $x(t)$ supplied by (I), is a sum of solutions of the form $x = (ct^j + O(t^{j-1}))e^{-\lambda t}$ ($\not\equiv 0$), as $t \to \infty$, where $Ac = \lambda c$, λ is an eigenvalue of A, c a corresponding eigenvector and j a non-negative integer. Hence, the partial sum $x = \Sigma(ct^j + O(t^{j-1}))e^{-\lambda t}$ of these solutions, belonging to the λ-values with the least real part and to the greatest j associated with these λ, is real and satisfies $x \geqq 0$ for large t, since $x = x(t)$ does. In

particular, λ is real and $c \geqq 0$. Hence, $x = ce^{-\lambda t}$ is non-negative (for all t) and is a solution of (17). But (16) and (17) imply that $x' = -\lambda ce^{-\lambda t}$ is non-positive (for all t); so that $\lambda \geqq 0$.

It follows that (I) can be considered as a generalization of the algebraic theorem of Perron ([6]; cf. Frobenius [2]) which states that a non-negative (constant) matrix A possesses at least one non-negative eigenvalue λ, corresponding to which there is a non-negative eigenvector c. (There also follows Perron's result which states that if $A > 0$, then, in the last assertion, $\lambda \geqq 0$ and $c \geqq 0$ can be improved to $\lambda > 0$ and $c > 0$. In fact, $\lambda > 0$ is needed, by Remark 1, to assure that $x(t) = ce^{-\lambda t} \to 0$ as $t \to \infty$, and $c > 0$ follows from the equations $c = \lambda^{-1}Ac$, since $A > 0$ and $(0 \neq)\ c \geqq 0$.)

Corresponding to the preceding deduction from (I), the assertion of (i) can be considered as a generalization of a particular case of Descartes's rule for the existence of a non-positive root for a polynomial.

(I) has the following corollary:

(I bis) *Let $A = A(t)$ be a completely monotone n by n matrix and let $p(t)$ be a positive vector having a completely monotone derivative on $0 < t < \infty$; that is, let*

$$(19) \qquad (-1)^j D^j A \geqq 0 \quad \text{for } j = 0, 1, \cdots$$

and

$$(20) \qquad p > 0 \text{ and } (-1)^j D^{j+1} p \geqq 0 \quad \text{for } j = 0, 1, \cdots,$$

where $0 < t < \infty$. Then

$$(21) \qquad px' = -Ax$$

has at least one solution vector $x = x(t) \not\equiv 0$ which is completely monotone on $0 < t < \infty$,

$$(22) \qquad (-1)^j D^j x \geqq 0 \quad \text{for } j = 0, 1, \cdots.$$

The same is true if $0 < t < \infty$ is replaced by $0 \leqq t < \infty$.

Remark. In [10], [7], linear differential equations with coefficients representable as Laplace transforms (6) *not* subject to $d\mu(s) \geqq 0$ were considered. Instead of the condition $d\mu(s) \geqq 0$ for all $s \geqq 0$, it was assumed that the contribution of small $s > 0$ to the integral (6) is small, in the sense that

$$\int_{+0} s^{-1} \mid d\mu(s) \mid < \infty;$$

for example, that $d\mu(s) \equiv 0$ for $0 \leqq s \leqq \epsilon$, where $\epsilon > 0$. The assertions of the theorems proved were to the effect that there exist non-trivial solutions which are (or that all solutions are) representable as Laplace transforms (6), where $d\mu \gtreqless 0$. The methods used were a "comparison of coefficients and a majorant method."

Theorems (i bis) and (I bis) seem to suggest that these theorems can be proved without the assumption for small $s (> 0)$ on the $d\mu(s)$ occurring in the representation of the coefficients, since it might be expected that the solutions supplied by (i bis) or (I bis) furnish a suitable majorant. Since, however, the proofs of (i bis) and (I bis) do not depend on a comparison of coefficients, but on qualitative arguments leading directly to (5) and (22), this is not the case. Simple examples illustrate this negative statement; for example, in the differential equation $x' = x/t$, where $0 < t < \infty$, the coefficient function $1/t$ is the Laplace transform (6) of $\mu(s) \equiv s$, but no solution $x = \text{const.}\, t \not\equiv 0$ is a Laplace transform.

3. *Proof of* (I) *and* (I bis). Let x^0 denote a vector satisfying $x^0 > 0$; for example, let $x^0 = (1, 1, \cdots, 1)$. For a positive integer l, let $x = x^l(t)$ denote the solution of (17) satisfying the initial condition $x(t) = x^0$ when $t = l$. The assumption (16) shows that any solution $x = x(t)$ of (17) satisfies $-x'(t) \geqq 0$ on any t-interval on which $x(t) \geqq 0$. Since $x^l(l) = x^0 > 0$, it follows that $x^l(t) > 0$ for t near l, hence $x^{l\prime}(t) \leqq 0$ for t near l. Thus $x^l(t) \geqq x^l(l) > 0$ for t less than and close to l. This argument shows that $x^l(t) > 0$ and $-x^{l\prime}(t) \geqq 0$ for $0 < t \leqq l$.

If $x^l = (x^l_1, \cdots, x^l_n)$, put $a_l = \max(x^l_1(1), \cdots, x^l_n(1))$; so that $a_l > 0$. Let $z^l(t) = x^l(t)/a_l$. Then $x = z^l(t)$ is a solution of (17) and satisfies

$$z^l(t) > 0 \text{ and } -z^{l\prime}(t) \leqq 0 \text{ for } 0 < t \leqq l \tag{23}$$

and, at $t = 1$, the components of the vector $z^l(t)$ satisfy

$$0 < z^l_k(1) \leqq 1 \text{ for } k = 1, 2, \cdots, n; \text{ while } z^l_{k(l)}(1) = 1 \tag{24}$$

holds for at least one index $k = k(l)$, where $1 \leqq k(l) \leqq n$.

It is clear from (24) that there exists a sequence of integers $l_1, l_2, \cdots$ with the property that

$$z^0 = \lim_{j\to\infty} z^l(1) \text{ exists} \qquad (l = l_j); \tag{25}$$

so that

$$z^0_k = 1 \text{ for some } k \qquad (1 \leqq k \leqq n). \tag{26}$$

The relations (25) and (26) imply that, if $x = x(t)$ is the solution (17) determined by the initial condition $x(1) = z^0$, then $x(t) \not\equiv 0$, and that $z^l(t) \rightarrow x(t)$ as $j \rightarrow \infty$, where $l = l_j$, holds uniformly on every closed, bounded subinterval of $0 < t < \infty$. Hence, the first inequality in (18) follows from (23), while the second is a consequence of the first and of (16), (17). This proves (I).

It will be clear from the remark following (I) and from the proof of (I bis) below that it is sufficient to consider the case in which the underlying t-range is the open half-line $0 < t < \infty$. Note that, since the k-th equation in the system (21) can be divided by $p_k(t) > 0$, assertion (I) implies that (21) has at least one solution $x = x(t) \not\equiv 0$ satisfying (18), which is equivalent to the cases $j = 0$ and $j = 1$ of (22). The set of all relations (22) will be proved by induction.

The equation (21) and the product rule for differentiation give

$$-pD^{j+1}x = \sum_{i=0}^{j-1} C_{ji}(D^{j-i}p)(D^{i+1}x) + \sum_{i=0}^{j} C_{ji}(D^{j-i}A)(D^i x),$$

where the $C_{jm} = j!/m!(j-m)!$ are the binomial coefficients. Hence, if $(-1)^m D^m x \geqq 0$ for $m = 0, 1, \cdots, j$, then (20) shows that

$$(-1)^j (D^{j-i}p)(D^{i+1}x) \geqq 0$$

for $i = 0, 1, \cdots, j-1$, while (19) implies that $(-1)^j(D^{j-i}A)(D^i x) \geqq 0$ for $i = 0, 1, \cdots, j$. Thus, from $p > 0$ and the last two formula lines,

$$(-1)^{j+1} D^{j+1} x \geqq 0.$$

This proves (I bis).

4. *Proof of* (i). In the proof of (i), it can be supposed that $f_0 \equiv 1$, for otherwise (1) can be divided by $f_0 > 0$. Then, if

$$g(t) = \exp \int_1^t f_1(s)\,ds > 0, \qquad (0 < t < \infty), \tag{28}$$

(1) can be written in the form

$$g^{-1}D(gD^{n-1}y) + \sum_{k=2}^{n} (-1)^{k-1} f_k(t) D^{n-k} y = 0. \tag{29}$$

Hence, if

$$x_1 = y, x_2 = -Dy, \cdots, x_{n-1} = (-1)^{n-2}D^{n-2}y, x_n = (-1)^{n-1}gD^{n-1}y, \tag{30}$$

(29) can be written as the system of linear, first order differential equations

$$(31) \qquad x_1' = -x_2, \cdots, x_{n-2}' = -x_{n-1}, x_{n-1}' = -x_n/g, x_n' = -g \sum_{k=1}^{n-1} f_{n-k+1} x_k.$$

In view of (2) and (28), condition (16) of (I) is satisfied when (31) is identified with (17). The definitions (30) and the assertion of (I) imply that (1) has a solution $y = y(t) \not\equiv 0$ satisfying the second of the inequalities (3) for $j = 0, 1, \cdots, n-1$. In particular, $y \geqq 0$ and $Dy \leqq 0$. Hence if $y = 0$ for some value of $t = t_0 > 0$, then $y(t) \equiv 0$ for $t \geqq t_0$, and therefore for $0 < t < \infty$. Since this is a contradiction, it follows that $y(t) > 0$ for $0 < t < \infty$. This proves (3).

The transition to the case of (i) in which $0 < t < \infty$ is replaced by $0 \leqq t < \infty$ is obvious. For in the latter situation, where the coefficients of (1) are defined and continuous on $0 \leqq t < \infty$, any solution $y = y(t)$ on $0 < t < \infty$ can be defined (by continuity) at $t = 0$ so as to be a solution on $0 \leqq t < \infty$. By continuity, the last set of inequalities in (3) holds at $t = 0$ also, while the first inequality in (3) follows at $t = 0$ from the monotony of y for $t > 0$.

5. *Proof of* (i bis). Assertion (I bis) implies (i bis) if (1) is written as the system of linear, first order differential equations

$$(32) \qquad x_1' = -x_2, \cdots, x_{n-1}' = -x_n, f_0 x_n' = -\sum_{k=1}^{n} f_{n-k+1} x_k.$$

In fact, if $x_1 = y$, $x_2 = -Dy, \cdots, x_n = (-1)^{n-1} D^{n-1} y$, then (32) can be identified with (21), where p is the vector $(1, \cdots, 1, f_0)$ and every element a_{ik} of the matrix $A(t)$ is either one of the functions $f_1, f_2, \cdots, f_n$ or is identically 0 or 1. Thus (4) implies conditions (19) and (20) of (I bis), and the assertion of (I bis) is equivalent to that of (i bis), by the definition of the vector $x = (x_1, \cdots, x_n)$.

6. The theorem (I) for differential equations has an analogue for difference equations:

(II) *Let $A = A(m) = (a_{ik}(m))$ be an n by n matrix function of the non-negative integer m. Suppose that*

$$(33) \qquad A(m) \geqq 0 \qquad (0 \leqq m < \infty),$$

and that, if I is the unit matrix, the reciprocal

$$(34) \qquad (I - A(m))^{-1} \text{ exists and is } \geqq 0 \qquad (0 \leqq m < \infty).$$

Then the linear, homogeneous, vector difference equation

$$(35)\qquad \Delta x(m) = -A(m)x(m)$$

has at least one solution $x = x(m)$ *satisfying*

$$(36)\qquad x(m) \geqq 0 \text{ and } -\Delta x(m) \geqq 0 \qquad (0 \leqq m < \infty),$$

and

$$(37)\qquad x(m) \neq 0 \qquad (0 \leqq m < \infty).$$

The relations (35) and (37) show that

$$(38)\qquad \Delta x(m) \neq 0 \text{ if } \det A(m) \neq 0.$$

It will be clear from the proof of (II) that if condition (34) is replaced by the assumptions that, for some integer $l \geqq 0$,

$$(34^*)\qquad (I - A(l))x = 0 \text{ has a solution, } 0 \neq x \geqq 0$$

and

$$(34^{**})\qquad (I - A(m))^{-1} \text{ exists and is } \geqq 0 \text{ for } 0 \leqq m < l,$$

then (II) remains valid if (37) is weakened to $x(m) \not\equiv 0$. In fact, (35) will then have a solution $x = x(m)$ satisfying (36), $x(m) \neq 0$ for $m = 0, 1, \cdots, l$, and $x(m) = 0$ for $m = l + 1, l + 2, \cdots$.

The difference analogue of (I bis) is as follows:

(II bis) *Let* $A(m)$ *satisfy the conditions of* (II). *In addition, let* $A(m)$ *be completely monotone, and let* $p = p(m)$ *be a positive vector having a completely monotone first difference; that is, let*

$$(39)\qquad (-1)^j \Delta^j A(m) \geqq 0 \text{ for } j = 0, 1, \cdots \qquad (0 \leqq m < \infty)$$

and

$$(40)\qquad p(m) > 0 \text{ and } (-1)^j \Delta^{j+1} p(m) \geqq 0 \qquad (0 \leqq m < \infty).$$

Then (35) *has at least one solution* $x = x(m)$ *which satisfies* (37) *and is completely monotone,*

$$(41)\qquad (-1)^j \Delta^j x(m) \geqq 0 \text{ for } j = 0, 1, \cdots \qquad (0 \leqq m < \infty).$$

7. *Proof of* (II) *and* (II bis). The equations (35) can be written in the form

$$(42)\qquad x(m+1) = x(m) - A(m)x(m) = (I - A(m))x(m),$$

which, since $I - A(m)$ is non-singular, is equivalent to

$$(43)\qquad x(m) = (I - A(m))^{-1}x(m+1).$$

The equations (42) and (43) show that if l is any non-negative integer and x^0 is an arbitrary vector, then (35) has a unique solution satisfying $x(l)=x^0$; in fact, (42) then determines $x(m)$ for every $m>l$, while (43) determines $x(m)$ for every $m<l$. Furthermore, (34) implies that, if $l>0$ and if the assigned initial condition $x(l)=x^0$ is a non-negative vector, then $x(m)\geqq 0$ for $0\leqq m\leqq l$, and so, by (35), $\Delta x(m)\leqq 0$ for $0\leqq m\leqq l$.

The proof of (II) can now be completed along the lines of the proof of (I), as follows: Let l be a positive integer, x^0 a positive vector, say $x^0=(1,1,\cdots,1)$, and $x^l=x^l(m)$ the solution of (35) satisfying $x^l(l)=x^0$. Then $x^l(m)\geqq 0$ and $-\Delta x^l(m)\geqq 0$ for $0\leqq m\leqq l$. In particular, $x^l(0)\geqq x^l(l)>0$, so that $a=a_l=\max(x^l{}_1(0),\cdots,x^l{}_n(0))$ is positive. Let $z^l=z^l(m)=x^l(m)/a_l$. Then $x=z^l(m)$ is a solution of (35) satisfying

$$z^l(m)>0 \text{ and } -\Delta z^l(m)\geqq 0 \text{ and } 0\leqq m\leqq l \tag{44}$$

and, at $m=0$, the components of z^l satisfy $0<z^l{}_k(0)\leqq 1$ for $k=1,\cdots,n$, while $z^l{}_{k(l)}(0)=1$ holds for at least one index $k=k(l)$, $0\leqq k(l)\leqq n$.

Let $l_1,l_2,\cdots$ be an increasing sequence of integers such that

$$z^0=\lim_{j\to\infty} z^l(0) \text{ exists} \qquad (l=l_j). \tag{45}$$

Clearly, $z^0{}_k=1$ for some k, where $1\leqq k\leqq n$. Let $x=x(m)$ be the solution of (35) determined by $x(0)=z^0$; in particular, $x(0)\neq 0$. If $x=x(m)$ in (42) is replaced by $x=z^l(m)$, it follows from (45) and an induction on m that

$$x(m)=\lim_{j\to\infty} z^l(m) \qquad (l=l_j) \tag{46}$$

for $m=0,1,\cdots$. Hence (36) follows from (44).

The remark made after (42), concerning the uniqueness of a solution of (35) satisfying a given initial condition, shows that (37) is satisfied, since $x(m)\not\equiv 0$. Hence (II) is proved.

The proof of (II bis) is similar to that of (I bis) and will be omitted; cf. [4], pp. 127-128.

8. *Proof of* (ii) *and* (ii bis). In view of the condition $f_0(m)>0$, it can be supposed that $f_0(m)\equiv 1$. Then (9) is

$$1-\sum_{k=1}^{n} f_k(m)>0; \tag{47}$$

in particular, $1-f_1(m)>0$. Put

$$g(0)=1 \text{ and } g(m)=g(m-1)/(1-f_1(m-1))>0 \quad \text{for } m=1,2,\cdots; \tag{48}$$

so that $\Delta g(m) = g(m+1)f_1(m)$ for $m = 0, 1, \cdots$ and $\Delta(g(m)\Delta^{n-1}y(m)) = g(m+1)(\Delta^n y(m) + f_1(m)\Delta^{n-1}y(m))$. Thus (7) can be written as

$$(49) \qquad g^{-1}(m+1)\Delta(g(m)\Delta^{n-1}y(m)) + \sum_{k=2}^{n}(-1)^{k-1}f_k(m)\Delta^{n-k}y(m) = 0.$$

In terms of the vector $x = x(m) = (x_1(m), \cdots, x_n(m))$ defined by

$$(50) \qquad x_1 = y, x_2 = -\Delta y, \cdots, x_{n-1} = (-1)^{n-2}\Delta^{n-2}y, x_n = (-1)^{n-1}g\Delta^{n-1}y,$$

the equation (49) is equivalent to the system of first order difference equations

$$(51) \qquad \Delta x_1 = -x_2, \cdots, \Delta x_{n-2} = -x_{n-1}, \qquad \Delta x_{n-1} = -x_n/g,$$

$$\Delta x_n(m) = -g(m+1)\sum_{k=1}^{n-1} f_{n-k+1}(m)x_k(m).$$

If (51) is identified with (35), it is seen that (33) holds.

In order to prove the existence of $(I - A)^{-1}$, let (51) be solved for $x(m)$, if possible. To this end, put

$$(52) \quad F_k(m) = f_k(m) + f_{k+1}(m) + \cdots + f_n(m) \geqq 0, \text{ where } k = 1, \cdots, n.$$

Multiply the k-th of the equations (51) by $g(n+1)F_{n-k+1}(m)$ or 1 according as $k = 1, \cdots, n-1$ or $k = n$, and add the resulting n equations. The result is

$$g(m+1)\sum_{k=1}^{n-1} F_{n-k+1}(m)x_k(m+1) + x_n(m+1)$$
$$= (1 - g(m+1)F_2(m)/g(m))x_n(m),$$

or since, $g(m+1)/g(m) = 1/(1 - f_1(m))$,

$$(53) \qquad x_n(m) = (1 - f_1(m))g(m+1)\sum_{k=1}^{n-1} F_{n-k+1}(m)x_k(m+1)/(1 - F_1(m)),$$

where $1 - F_1(m) > 0$, by (47). It follows that $x_{n-1}(m), x_{n-2}(m), \cdots, x_1(m)$ can successively be expressed in terms of the components of $x(m+1)$, by using (53) and the $(n-1)$-st, the $(n-2)$-nd, $\cdots$ of the equations (51). For the first $n-1$ of the equations (51) can be written (in the reverse order) as

$$(54) \qquad x_{n-1}(m) = x_{n-1}(m+1) + x_n(m)/g(m),$$

$$x_{n-2}(m) = x_{n-2}(m+1) + x_{n-1}(m), \cdots, x_1(m) = x_1(m+1) + x_2(m).$$

Since the system (53)-(54) is equivalent to (43), it follows that $(I - A)^{-1}$ exists, in view of (47), and that $(I - A(m))^{-1} \geqq 0$, by (47) and (52).

Thus (II) is applicable to the system (51). Hence there exists a

solution $x = x(m)$ satisfying (36) and (37). It follows by (50) that (7) possesses a solution $y = y(m) \not\equiv 0$ such that (10) holds if the $>$ is replaced by $\geqq$ in the first inequality in (10). However, if $y(l) = 0$ for some $l (\geqq 0)$, then $y(m) \equiv 0$ for $m \geqq l$, since $y(m) \geqq 0$ and $-\Delta y(m) \leqq 0$. In this case, $x(m) \equiv 0$ for $m \geqq l$, by (50). Since this contradicts (37), it follows that $y(m) > 0$ for $m = 0, 1, \cdots$, and so (ii) is proved.

As to the Remark 2, following the statement of (ii), note that $\det A(m) = (-1)^n g(m+1) f_n(m)$, by (51); so that $\det A(m) = 0$ if and only if $f_n(m) = 0$. Thus, the assumption of Remark 2 and (38) show that "$\Delta x(m) \equiv 0$ for $m \geqq l$" cannot hold for any l. But if $n > 1$ and $\Delta y(m) = 0$ for $m = l$, then $\Delta y(m) = 0$, and so $\Delta x(m) \equiv 0$ for $m \geqq l$, by (50). Hence

$$-\Delta y(m) > 0 \quad \text{for} \quad m = 0, 1, \cdots.$$

As to Remark 3, it is sufficient to apply the comments made above on (34*), (34**) and (II). The proof of (ii) shows that $(I - A(m))^{-1}$ exists (and is $\geqq 0$) for a given m, if and only if the inequality in (9) holds. For some m, let

$$f_0(m) - \sum_{k=1}^{n} f_k(m) = 0. \tag{55}$$

Using the notation of the proof of (ii), where $f_0(m) \equiv 1$ and where (51) is equivalent to $\Delta x(m) = -A(m)x(m)$, it is seen that $x = (1/g(m), \cdots, 1/g(m), 1)$ is a solution of $A(m)x = x$, if it is recalled that

$$g(m+1)/g(m) = 1/(1 - f_1(m))$$

and $1 - f_1(m) = F_2(m) + \cdots + F_n(m)$. Hence, if (9) is weakened to allow $\geqq$ in place of $>$, then (34*) and (34**) hold for the least $m = l \geqq 0$ satisfying (55).

The assertion (ii bis) follows from (II bis) in exactly the same way as (i bis) does from (I bis). The proof of (ii bis) will therefore be omitted.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] O. Dunkel, "Regular singular points of a system of homogeneous linear differential equations of the first order," *Proceedings of the American Academy of Arts and Sciences*, vol. 38 (1902-03), pp. 341-370.

[2] G. Frobenius, "Ueber Matrizen aus positiven Elementen," *Sitzungsberichte der Kgl. Preussischen Akademie der Wissenschaften zu Berlin*, 1908, pp. 471-476 and 1909, pp. 514-518; "Ueber Matrizen aus nicht negativen Elementen," *ibid.*, 1912, pp. 456-477.

[3] P. Hartman and A. Wintner, "On the Laplace-Fourier transcendents," *American Journal of Mathematics*, vol. 71 (1949), pp. 367-372.

[4] ——— and A. Wintner, "On linear difference equations of second order," *ibid.*, vol. 72 (1950), pp. 124-128; cf. p. 624.

[5] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[6] O. Perron, "Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus," *Mathematische Annalen*, vol. 64 (1907), pp. 1-76.

[7] C. R. Putnam and A. Wintner, "Linear differential equations with almost periodic or Laplace transform coefficients," *American Journal of Mathematics*, vol. 73 (1951), pp. 792-806.

[8] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *ibid.*, vol. 69 (1947), pp. 87-89.

[9] ———, "Asymptotic integrations of the adiabatic oscillator," *ibid.*, vol. 69 (1947), pp. 251-272.

[10] ———, "On the small divisors in integrations by Laplace transforms," *ibid.*, vol. 73 (1951), pp. 173-180.

A CONNECTION BETWEEN THE WHITEHEAD AND THE PONTRYAGIN PRODUCT.* [1]

By Hans Samelson.

1. Introduction. Let X be a 1-connected (i. e. arcwise and simply connected) topological space, and let Ω be the space of loops (closed paths) in X, with base point x_0 (cf. [5] for this concept). We recall some definitions and facts.

(a) If α and β are elements of the homotopy groups $\pi_{p+1}(X)$, $\pi_{q+1}(X)$, we denote, as customary, by $[\alpha, \beta]$ their Whitehead product (cf. e. g. [8]); it is an element of $\pi_{p+q+1}(X)$.

(b) The space Ω possesses a natural multiplication (composition of loops, as used in the definition of fundamental group), and this gives rise to the Pontryagin product for the (singular) homology group of Ω; if a, b are elements of $H_r(\Omega)$, $H_s(\Omega)$, then the product $a * b$ belongs to $H_{r+s}(\Omega)$ (cf. [2] for definitions and algebraic properties).

(c) There is a natural isomorphism T between $\pi_n(X)$ and $\pi_{n-1}(\Omega)$ (actually there are several such, cf. § 3); we let h denote the standard map, introduced by Hurewicz, of the homotopy groups of a space into the homology groups, and define $\tau: \pi_n(X) \to H_{n-1}(\Omega)$ by $\tau = h \circ T$. (T and τ are related to transgression, cf. [7], p. 452.) We can now state the result of the present note, with T meaning $\delta \circ p^{-1}$, cf. §§ 2, 3.

THEOREM. *If* $\alpha \in \pi_{p+1}(X)$, $\beta \in \pi_{q+1}(X)$, p, $q \geqq 1$, then

$$\tau[\alpha, \beta] = (-1)^p(\tau\alpha * \tau\beta - (-1)^{pq}\tau\beta * \tau\alpha).$$

The sign $(-1)^p$ depends of course on the choice of the map T. We remark that special cases of the formula have been known to Hurewicz, Serre, G. W. Whitehead, and also acknowledge conversations with J. C. Moore and J. Dugundji, in which the problem was raised.

We give two proofs; the first proof gives the formula as consequence of some general facts, but leaves the sign open; the second proof is very elementary and direct. In § 2 we derive a homomorphism for pairs of fiber

* Received April 29, 1953.

[1] The work on this paper was performed under OOR contract no. DA-36-034-ORD-975RD.

spaces, which seems of some interest, although the special case needed later can be given a much simpler treatment.

2. A fiber space relation. Let $p: E \to B$ be a fiber map in the sense of Serre [7], with 0-connected total space E, base B and fibers $F_x = p^{-1}(x)$. Suppose a subset E' of E is a fiber space over $p(E') = B'$ relative to $p' = p \mid E$ (p restricted to E'), with E', B' and the fibers $F'_x = p'^{-1}(x)$ again 0-connected. Let x_0 be a point of B', and put $F = F_{x_0}$, $F' = F'_{x_0}$. Clearly $F' = E' \cap F$. It is well known that the map p induces an isomorphism between the homotopy groups $\pi_n(E, F)$ and $\pi_n(B, x_0) = \pi_n(B)$ (and similarly for p') ; the proof given in [6], p. 90, applies in the present case. The symbols $\pi_1(E, F)$, $\pi_0(B)$ etc. denote sets without group structure (cf. [1], p. 167) ; p is still $1:1$, onto in dimension 1.

We denote by T ("transgression") the composition $\delta \circ p^{-1}$ of the isomorphism $p^{-1}: \pi_n(B) \to \pi_n(E, F)$ and the boundary map $\delta: \pi_n(E, F) \to \pi_{n-1}(F)$; we have the analogous T for E'. We now construct a similar map for the relative groups. The spaces (E, E', F) form a triad (in the sense of [1]). We consider the following diagram (it applies to an arbitrary triad) :

$$
\begin{tikzcd}
& {} \arrow[d] & {} \arrow[d] & & {} \arrow[d] & & {} \arrow[d] & \\
{} \arrow[r] & \pi_n(E') \arrow[r] \arrow[dd] & \pi_n(E) \arrow[rr] \arrow[dd] \arrow[dr] & & \pi_n(E, E') \arrow[rr] \arrow[dd] & & \pi_{n-1}(E') \arrow[r] \arrow[dd] & {} \\
& & & \pi_n(E, F') \arrow[ur] \arrow[dl] & & & & \\
{} \arrow[r] & \pi_n(E', F') \arrow[r] \arrow[dd] & \pi_n(E, F) \arrow[rr] \arrow[dd] & & \pi_n(E; E', F) \arrow[rr] \arrow[dd] & & \pi_{n-1}(E', F') \arrow[r] \arrow[dd] \arrow[dl] & {} \\
& & & & & \pi_{n-1}(E, F') \arrow[dr] & & \\
{} \arrow[r] & \pi_{n-1}(F') \arrow[r] \arrow[d] & \pi_{n-1}(F) \arrow[rr] \arrow[d] & & \pi_{n-1}(F, F') \arrow[rr] \arrow[ur] \arrow[d] & & \pi_{n-1}(F') \arrow[r] \arrow[d] & {} \\
& {} & {} & & {} & & {} &
\end{tikzcd}
$$

DIAGRAM 1.

All maps are the standard maps; commutativity clearly holds everywhere except that for $n > 2$ the two maps from $\pi_n(E; E', F)$ to $\pi_{n-1}(E, F')$ are negatives of each other (cf. Lemma 3. 5. 5 of [1], p. 177 for the proof). On the other hand, the map p maps the triad $(E; E', F)$ into the triad $(B; B', x_0)$ and therefore it maps the middle row of Diagram 1 into the homotopy sequence of (B, B') (i. e. the commutativity relations hold) ; in view of the isomorphisms noted above and the "five lemma" ([2], p. 16) this is actually an isomorphism of the two sequences ($\pi_2(E; E', F)$ has to be treated separately). We compose the inverse p^{-1} of this isomorphism with the maps δ from the middle row to the bottom row of Diagram 1, and put $T = \delta \circ p^{-1}$ ("transgression"). The map T can be made into a homo-

morphism of the respective sequences, i. e. commutativity can be introduced, by changing the sign in every other dimension.

PROPOSITION 1. *There exists a homomorphism T_0 of the homotopy sequence of (B, B') into that of (F, F') of degree -1, i. e. which lowers dimension by one; the maps of $\pi_n(B')$ into $\pi_{n-1}(F')$, resp. of $\pi_n(B)$ into $\pi_{n-1}(F)$ are given by $(-1)^n\delta \circ p'^{-1}$, resp. $(-1)^n\delta \circ p^{-1}$. (We recall that $\pi_1(F, F')$ has no group structure; $\pi_0(F')$ and $\pi_0(F)$ reduce to the neutral element.)*

COROLLARY. *If E' and E are contractible, then the homotopy sequences of (B, B') and (F, F') are isomorphic with a shift of 1 in dimension (for dimension 1 this reduces to the fact that $\pi_1(B')$, $\pi_1(B)$, $\pi_0(F')$, $\pi_0(F)$ are all trivial).*

Proof of the Corollary. The maps δ are now isomorphisms, as seen from Diagram 1.

Proposition 1 applies, appropriately interpreted, to the slightly more general case of a fiber map f of a fiber space (E', B', p') into another (E, B, p) (i. e. a pair of maps $f_e : E' \to E$, $f_b : B' \to B$, such that $f_b \circ p' = p \circ f_e$). The relative groups $\pi_n(B, B')$, $\pi_n(F, F')$ have to be understood as the relative groups of the corresponding maps, i. e. the relative groups of the mapping cylinder modulo the mapped space; similarly for the triad groups $\pi_n(E; E', F)$. The proof is essentially the same as before, modified by the introduction of the appropriate mapping cylinders.

3. A special case. Let (X, Y) be a pair, with both spaces 1-connected; take a point x_0 in Y. Let E' be the space of paths in Y, which end at x_0 (cf. [1], [7]); it is a fiber space over Y, with projection p' (p' maps each path on its initial point), and fiber Ω_Y (space of loops in Y, based at x_0); let E, p, Ω_X be the corresponding objects for X. Then E' is contained naturally in E, and p' is the restriction of p to E', so that we are in the situation of § 2, with Ω_Y, Ω_X playing the roles of F', F. Since E and E' are contractible ([7], p. 471), we can apply the corollary of § 2, and obtain an isomorphism T_0 of degree -1, of the homotopy seqtence of (X, Y) onto that of (Ω_X, Ω_Y).

Actually this can be established in a much simpler way: Let i and n be integers with $1 < i \leqq n$. If f is a map from the n-cube I^n (product of n copies of the interval $I = [0, 1]$) to X, with the boundary $\dot{I}^n$ going into x_0, we define T_if as the map from I^{n-1} to Ω_X given by

$$T_if(x_1, \cdots, x_{n-1})(t) = f(x_1, \cdots, x_{i-1}, t, x_i, \cdots, x_{n-1})$$

(the definition makes sense, if only the faces $x_i = 0$ and $x_i = 1$ of I^n go into x_0). Clearly the operator T_i induces an isomorphism, also called T_i, between $\pi_n(X) = \pi_n(X, x_0)$ and $\pi_{n-1}(\Omega_X)$ (with base point e_0, where e_0 is the degenerate path $(e_0(t) = x_0$ for $0 \leqq t \leqq 1))$; one could even allow $i = 1$, replacing addition in $\pi_{n-1}(\Omega_X)$ by the multiplication in Ω_X mentioned in § 1, (b). For $i < n$ the operator T_i defines an isomorphism of the relative groups $\pi_n(X, Y)$ and $\pi_{n-1}(\Omega_X, \Omega_Y)$ (this is, for $i = n - 1$, a special case of a remark in [4], p. 493), and commutes with the boundary operator; in particular T_2 maps the homotopy sequence of (X, Y) isomorphically onto that of (Ω_X, Ω_Y), with degree -1. (For the group $\pi_2(X, Y)$ one has to replace T_2 by $-T_1$; it is then mapped $1:1$ onto $\pi_1(\Omega_X, \Omega_Y)$.) One verifies that for $1 \leqq i, j \leqq n$ the relations $T_i = (-1)^{i+j} T_j$ hold; this follows from the fact that the interchange of two axes is an orientation reversing homeomorphism of I^n. The operation T_n, applied to $\pi_n(X)$, coincides with the map $\delta \circ p^{-1}$ considered in § 2: If f is a map of $(I^n, \dot{I}^n)$ into (X, x_0), we define a map f' of I^n into E by $f'(x_1, \cdots, x_n)(t) = f(x_1, \cdots, x_{n-1}, t + x_n(1 - t))$. Clearly $p \circ f' = f$, and if the point $(x_1, \cdots, x_n)$ belongs to $\dot{I}^n - I^{n-1}$, then $f'(x_1, \cdots, x_n) = e_0$; so that f' defines an element of $\pi_n(E, \Omega_X)$, projecting into the element of $\pi_n(X)$ defined by f. Applying δ, i. e. putting $x_n = 0$, we obtain a map f'' of $(I^{n-1}, \dot{I}^{n-1})$ into (Ω_X, e_0), given by $f''(x_1, \cdots, x_{n-1})(t) = f(x_1, \cdots, x_{n-1}, t)$, which is exactly $T_n f$.

4. An application. We apply § 3 to the case where X is the cartesian product of two (oriented) spheres S^{p+1}, S^{q+1} of dimensions $p+1, q+1$ $(p, q \geqq 1)$, and where Y is the "union" $S^{p+1} \vee S^{q+1}$, i. e. $S^{p+1} \times z_0 \cup y_0 \times S^{q+1}$, with $y_0 \varepsilon S^{p+1}$, $z_0 \varepsilon S^{q+1}$. We recall some known facts [8]. The Künneth formula (for a proof cf. [2]), applied to the well-known homology groups of spheres, shows that the first non-vanishing relative homology group of $X \bmod Y$ occurs in dimension $p + q + 2$, and that this group is infinite cyclic; by the relative Hurewicz theorem [5] the same situation holds in homotopy. Further, $\pi_n(Y)$ maps onto $\pi_n(X)$ in all dimensions, so that, by exactness, δ maps $\pi_{p+q+2}(X, Y)$ isomorphically into $\pi_{p+q+1}(Y)$; i. e. the kernel of the map $\pi_{p+q+1}(Y) \to \pi_{p+q+1}(X)$ is infinite cyclic; if α_0, resp. β_0 are the elements of $\pi_{p+1}(S^{p+1})$, resp. $\pi_{q+1}(S^{q+1})$, defined by the identity maps, and α_1, β_1 are their images, under inclusion, in Y, the Whitehead product $[\alpha_1, \beta_1]$ is a generator of this kernel (this is implicit in the definition of $[\ ,\ ]$, as formulated in [8], p. 201). We now apply $T_0 (= T_2)$; it follows that the first non-vanishing relative homotopy and homology group of (Ω_X, Ω_Y) occur in dimension $p + q + 1$, and that they are isomorphic and infinite cyclic.

(For p or $q=1$ one has to use here the fact, that Ω_X, as H-space, is simple, [7], p. 479.) We have the diagram

$$\begin{tikzcd}
\arrow[r] & \pi_{p+q+1}(\Omega_X, \Omega_Y) \arrow[r, "\delta"] \arrow[d, leftrightarrow] & \pi_{p+q}(\Omega_Y) \arrow[r] \arrow[d, "h"] & \pi_{p+q}(\Omega_X) \arrow[r] \arrow[d] & {} \\
\arrow[r] & H_{p+q+1}(\Omega_X, \Omega_Y) \arrow[r, "\delta"] & H_{p+q}(\Omega_Y) \arrow[r, "i"] & H_{p+q}(\Omega_X) \arrow[r] & {}
\end{tikzcd}$$

DIAGRAM 2.

The image of δ in the upper line is the infinite cyclic group generated by $T[\alpha_1, \beta_1]$. We now determine the image of δ in the lower line, i. e. the kernel of i.

5. The Pontryagin ring. We recall the concept of Pontryagin ring, defined for the space Ω of loops in any topological space: the natural composition or multiplication of loops (denoted by $\cdot$) can be considered as a map γ from $\Omega \times \Omega$ to Ω, and induces a multiplication (the Pontryagin product) in the homology group of Ω (cf. [2] for details). The Pontryagin product $a * b$ of two homology classes of Ω is the image $\gamma(a \otimes b)$ of the element $a \otimes b$ of $H(\Omega \times \Omega)$.

The Pontryagin rings of the loop spaces Ω_1, Ω_2 of S^{p+1}, S^{q+1}, are polynomial rings in the variables $\tau\alpha_0$, $\tau\beta_0$ (of dimensions p, q) (cf. [2], II. 1. 3 or [9], p. 215). The space Ω_X of our present X is clearly the cartesian product of Ω_1 and Ω_2; this implies that $H(\Omega_X) = H(\Omega_1) \otimes H(\Omega_2)$, since $H(\Omega_1)$ and $H(\Omega_2)$ are free groups. Moreover, Ω_X is the direct product of Ω_1 and Ω_2 with respect to the multiplication γ. It follows easily that the Pontryagin ring of Ω_X is the (skew) tensor product of those of Ω_1 and Ω_2, i. e. that multiplication satisfies the rule

$$(a \otimes b) * (c \otimes d) = (-1)^{rs}(a * c) \otimes (b * d),$$

with $r = \dim b$, $s = \dim c$. (One shows, that for the map $\lambda\colon \Omega_1 \times \Omega_2 \to \Omega_2 \times \Omega_1$, given by $\lambda(x, y) = (y, x)$, one has $\lambda(c \otimes b) = (-1)^{rs} b \otimes c$; the desired property of $*$ follows then from the commutative diagram

$$\begin{tikzcd}
(\Omega_1 \times \Omega_2) \times (\Omega_1 \times \Omega_2) \arrow[r, "\gamma"] \arrow[d, leftrightarrow] & \Omega_1 \overset{\bullet}{\times} \Omega_2 \arrow[d, leftrightarrow] \\
(\Omega_1 \times \Omega_1) \times (\Omega_2 \times \Omega_2) \arrow[r, "\gamma \times \gamma"] & \Omega_1 \times \Omega_2 .)
\end{tikzcd}$$

DIAGRAM 3.

We can identify $\tau\alpha_0$, $\tau\beta_0$ with the elements $\tau\alpha_0 \otimes e_0$, $e_0 \otimes \tau\beta_0$ in $H_*(\Omega_X)$

$= H_*(\Omega_1) \otimes H_*(\Omega_2)$, so that $H_*(\Omega_X)$ is the ring (with unit e_0) generated by $\tau\alpha_0$, $\tau\beta_0$, subject to the relation

$$\tau\alpha_0 * \tau\beta_0 = (-1)^{pq}\tau\beta_0 * \tau\alpha_0 \ (= \tau\alpha_0 \otimes \tau\beta_0).$$

On the other hand, it has been shown in [2], III. 1. B., that the Pontryagin ring $H_*(\Omega_Y)$ of the present Y is the free associative algebra (with unit e_0) generated by the two elements $\tau\alpha_1$, $\tau\beta_1$, the (spherical) homology elements determined by $T\alpha_1$, $T\beta_1$. The inclusion $i\colon \Omega_Y \subset \Omega_X$ induces a homomorphism of the Pontryagin rings, since it is homomorphic with respect to γ, and clearly the elements $\tau\alpha_1$, $\tau\beta_1$ are mapped into $\tau\alpha_0$, $\tau\beta_0$. It is algebraically obvious that in dimension $p+q$ the kernel of i is the infinite cyclic group generated by $\tau\alpha_1 * \tau\beta_1 - (-1)^{pq}\tau\beta_1 * \tau\alpha_1$. Our main result, for the space Y under consideration, now follows immediately from Diagram 2: Since the groups on the left are isomorphic, h must map the generator $T[\alpha_1, \beta_1]$ of the infinite cyclic group $\delta\pi_{p+q+1}(\Omega_X, \Omega_Y)$ onto $\pm$ the generator

$$\tau\alpha_1 * \tau\beta_1 - (-1)^{pq}\tau\beta_1 * \tau\alpha_1$$

of the infinite cyclic group $\delta H_{p+q+1}(\Omega_X, \Omega_Y)$.

6. The general case. Let now X stand for an arbitrary 1-connected space, and let α, β be elements of $\pi_{p+1}(X)$, $\pi_{q+1}(X)$. We represent α and β by maps of S^{p+1} and S^{q+1} into X, and construct the obvious map f of $S^{p+1} \vee S^{q+1}$ into X. Then the elements α_1, β_1 of § 4 map into α, β under f. There is an induced map f' of the loop space of $S^{p+1} \vee S^{q+1}$ into that of X; f and f' are homomorphic with respect to $[\ ,\ ]$ and $*$; f' commutes with h; and we have $f' \circ T = T \circ f$. The result for X now follows by applying f' to the result for $S^{p+1} \vee S^{q+1}$, which holds by § 5.

7. Second proof. Our second proof is based on an interpretation of the Whitehead product, which goes back to Hurewicz and G. W. Whitehead; we present a derivation of this interpretation. X is again a 1-connected space, x_0 a point in it. We recall that in Ω_X an inversion σ, $(x \to x^{-1})$ is defined (by replacing the parameter t of any loop by $1-t$); clearly $\sigma^2 = 1$; the map $x \to x \cdot x^{-1}$, i. e. $\gamma \circ 1 \times \sigma \circ \Delta$, where Δ is the diagonal map of Ω_X into $\Omega_X \times \Omega_X$, is null-homotopic ([9], p. 210).

Let f and g be maps of $(I^{p+1}, \dot{I}^{p+1})$ and $(I^{q+1}, \dot{I}^{q+1})$ into (X, x_0), representing elements α and β of the respective homotopy groups. Then $[\alpha, \beta]$ is given by the map k of

$$S = S^{p+q+1} = \dot{I}^{p+q+2} = (I^{p+1} \times I^{q+1})^{\cdot} = I^{p+1} \times \dot{I}^{q+1} \cup \dot{I}^{p+1} \times I^{q+1},$$

defined by $k(x, y) = f(x)$, if $y \in \dot{I}^{q+1}$, $= g(y)$, if $x \in \dot{I}^{p+1}$; the base point on

S is $\omega=(0,\cdots,0)$. We write $I^{p+1}=I^p\times I$, $I^{q+1}=\dot{I}^q\times I$, and $I^{p+q+2}=I^p\times I\times I^q\times I$. Let K be the subset of S, given by

$$I^p\times I\times \dot{I}^q\times I\cup \dot{I}^p\times I\times I^q\times I\cup I^p\times 0\times I^q\times 0;$$

by collapsing the two factors I in the first two sets in this union one can contract K over itself into $I^p\times 0\times I^q\times 0$, and then into ω, with ω stationary. We now construct a map ϕ of $E=I^{p+q+1}=I^p\times I^q\times I$ into S as follows: for any $x\varepsilon I^p$, $y\varepsilon I^q$ the interval $x\times y\times I$ is divided into 6 parts by the t-values $0,\frac{1}{4},\frac{3}{8},\frac{1}{2},\frac{5}{8},\frac{3}{4},1$ and mapped in the obvious piecewise linear fashion on the closed polygon in S with successive vertices

$$\omega,\ (x,0,y,0),\ (x,1,y,0),\ (x,1,y,1),\ (x,0,y,1)\ (x,0,y,0),\ \omega.$$

If either $x\varepsilon \dot{I}^p$ or $y\varepsilon \dot{I}^q$ or $t=0$ or $t=1$, then $\phi(x,y,t)\varepsilon K$; so that $\phi(\dot{E})\subset K$. For $x\varepsilon I^p-\dot{I}^p$, $y\varepsilon I^q-\dot{I}^q$, $\frac{1}{4}<t<\frac{3}{8}$ one verifies that (x,y,t) is the only point of E mapped onto $(x,8t-2,y,0)$ by ϕ, that ϕ is locally $1:1$ in the neighborhood of (x,y,t) and that the local degree, using the natural orientations of E and S, is $(-1)^p$. It follows that the generator ϵ of $\pi_{p+q+1}(E,\dot{E})$, represented by the identity map of E, is mapped by ϕ into $(-1)^p\eta$, where η is the generator of $\pi_{p+q+1}(S,K)$ represented by the identity map of S.

Put $F=I^p\times I^q$ so that $E=F\times I$. Let s_1, s_2 be the customary maps of I^p, I^q onto S^p, S^q by collapsing the boundary to a point; let $s=s_1\times s_2$ the induced map of $(F,\dot{F})$ onto $(S^p\times S^q, S^p\vee S^q)$. Let $T=T_{p+q+1}$, as defined in § 3, so that $T(k\circ\phi)(x,y)(t)=k\circ\phi(x,y,t)$. One verifies that $T(k\circ\phi)$ can be factored in the form $c\circ s$ (if $x\varepsilon\dot{I}^p$, then $T(k\circ\phi)(x,y)$ depends on y only, if $y\varepsilon\dot{I}^q$, then $T(k\circ\phi)(x,y)$ depends on x only); here c is defined as follows: the maps $T_{p+1}f$, $T_{q+1}g$ can be factored into $f'\circ s_1$, $g'\circ s_2$, with f', g' mapping S^p, S^q into Ω_X; then $c(x,y)$ is

$$(e\cdot(f'(x)\cdot g'(y))\cdot((f'(x)^{-1}\cdot g'(y)^{-1})\cdot e).$$

Clearly c is homotopic to the map d, defined by

$$d(x,y)=(f'(x)\cdot g'(y))\cdot(f'(x)^{-1}\cdot g'(y)^{-1}).$$

We now contract K over itself to ω, as indicated above, and extend this to a deformation ψ_t of S, with $\psi_0=$ identity. One verifies that for each t the map $T(k\circ\psi_t\circ\phi)$ can be factored in the form $c_t\circ s$; the c_t consequently form a homotopy of $c=c_0$. Clearly $\psi_1\circ\phi$ maps ϵ into $(-1)^p\eta$, just as ϕ did. But $\psi_1\circ\phi(\dot{E})=\omega$, so that $\psi_1\circ\phi$ maps $(E,\dot{E})$ with degree $(-1)^p$ into (S,ω); it follows that $k\circ\psi_1\circ\phi$ represents $(-1)^p[\alpha,\beta]$. By definition then the map $T(k\circ\psi_1\circ\phi)$ maps the natural generator of $H_{p+q}(F,\dot{F})$ into $(-1)^p\tau[\alpha,\beta]$ (we identify here $H_{p+q}(\Omega_X)$ and $H_{p+q}(\Omega_X,e_0)$). Since this

map factors into $c_1 \circ s$, it follows that c_1, and therefore also c and d, map the natural generator of $H_{p+q}(S^p \times S^q)$ into $(-1)^p\tau[\alpha, \beta]$; we have used the fact that s maps $H_r(F, F)$ isomorphically onto $H_r(S^p \times S^q, S^p \vee S^q)$ (cf. [3], p. 266), that the map from $H_{p+q}(S^p \times S^q)$ to the relative group in the homology sequence of $(S^p \times S^q, S^p \vee S^q)$ is an isomorphism onto, and that c_1 maps $S^p \vee S^q$ into the point e_0. Our problem is now reduced to the discussion of the homology type of d.

8. The homology type of d. Let δ_1, δ_2 be the diagonal maps of S^p into $S^p \times S^p$, resp. S^q into $S^q \times S^q$; let λ be the permutation map of $S^p \times S^q$ onto $S^q \times S^p$; σ and γ are still inversion and multiplication in Ω_X. Then d can be written as the composition of the following maps: $\delta_1 \times \delta_2$, $1 \times \lambda \times 1$, $f' \times g' \times f' \times g'$, $1 \times 1 \times \sigma \times \sigma$, $\gamma \times \gamma$, γ. Let a, b denote the natural generators of $H_p(S^p)$, $H_q(S^q)$, so that by definition $f'(a) = \tau\alpha$, $f'(b) = \tau\beta$; we have to determine $d(a \otimes b)$. We use e as a generic symbol for the homology class of a point. It is well known that

$$\delta_1(a) = a \otimes e + e \otimes a, \text{ and } \delta_2(b) = b \otimes e + e \otimes b,$$

so that

$$\begin{aligned}&\delta_1 \times \delta_2(a \otimes b)\\ &\quad = a \otimes e \otimes b \otimes e + a \otimes e \otimes e \otimes b + e \otimes a \otimes b \otimes e + e \otimes a \otimes e \otimes b.\end{aligned}$$

Applying a remark of § 5, we have $\lambda(a \otimes b) = (-1)^{pq} b \otimes a$, $\lambda(a \otimes e) = e \otimes a$, $\lambda(e \otimes b) = b \otimes e$. As for σ, we note that, in consequence of a remark at the beginning of § 7, the map $\gamma \circ 1 \times \sigma \circ f' \times f' \circ \delta_1 = \gamma \circ 1 \times \sigma \circ \Delta \circ f'$ is homotopic to 0. Using the equation $\delta_1(a) = a \otimes e + e \otimes a$ and the fact that e is unit element in the Pontryagin ring of Ω_X, one finds $f'(a) + \sigma \circ f'(a) = 0$. If one now applies the factors of the map d in succession to $a \otimes b$ and makes use of the relations just stated, one obtains the result

$$d(a \otimes b) = \tau\alpha * \tau\beta - (-1)^{pq}\tau\beta * \tau\alpha.$$

As remarked at the end of § 7, we have $d(a \otimes b) = (-1)^p\tau[\alpha, \beta]$, and our main result is established.

9. Remarks.

1. The relation $[\alpha, \beta] = (-1)^{(p+1)(q+1)}[\beta, \alpha]$ is known to hold [8]; an easy computation shows that the right side of our main relation, with the sign as determined, satisfies the same relation.

2. If X is an n-sphere, and $\alpha = \beta = \iota_n$ are the elements of $\pi_n(S^n)$ determined by the identity map, we have

$$\tau[\iota_n, \iota_n] = \begin{matrix} -2\tau\iota_n * \tau\iota_n, & n \text{ even} \\ 0 \quad , & n \text{ odd.} \end{matrix}$$

This expresses the fact that the Hopf invariant of $[\iota_n, \iota_n]$ is ± 2 or 0, depending on the parity of n.

3. If $\gamma \varepsilon \pi_{r+1}(X)$, with α and β as before, then one verifies that the τ-image of

$$[\alpha, \beta, \gamma] = (-1)^{(p+1)r}[\alpha, [\beta, \gamma]] + (-1)^{(q+1)p}[\beta, [\gamma, \alpha]] + (-1)^{(r+1)q}[\gamma, [\alpha, \beta]]$$

vanishes. It seems a reasonable conjecture that the homotopy element $[\alpha, \beta, \gamma]$ itself vanishes; the validity of this Jacobi identity remains an open question.

4. The computation of § 8 is valid in any space, which possesses a continuous multiplication with a homotopy-unit and an inversion (such as γ, e_0, σ for Ω_X).

THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] A. L. Blakers and W. S. Massey, "The homotopy groups of a triad I," *Annals of Mathematics*, vol. 53 (1951), pp. 161-205.

[2] R. Bott and H. Samelson, "On the Pontryagin product in spaces of paths," to appear in *Comm. Math. Helv.*

[3] S. Eilenberg and N. E. Steenrod, *Roundations of algebraic topology*, Princeton, 1952.

[4] R. H. Fox, "Homotopy and torus homotopy groups," *Annals of Mathematics*, vol. 49 (1948), pp. 471-510.

[5] S. T. Hu, "An exposition of the relative homotopy theory," *Duke Mathematical Journal*, vol. 14 (1947), pp. 991-1033.

[6] N. E. Steenrod, *The topology of fiber bundles*, Princeton, 1951.

[7] J. P. Serre, "Homologie singulière des espaces fibrés, applications," *Annals of Mathematics*, vol. 54 (1951), pp. 425-505.

[8] G. W. Whitehead, "A generalization of the Hopf invariant," *Annals of Mathematics*, vol. 51 (1950), pp. 192-237.

[9] ———, "On the Freudenthal theorems," *Annals of Mathematics*, vol. 57 (1953), pp. 209-228.

GALOIS THEORY OF DIFFERENTIAL FIELDS.*

By E. R. Kolchin.[1]

Table of Contents.

---

* Received April 21, 1953. This paper was an address delivered before the New York meeting of the American Mathematical Society on April 26, 1952, by the invitation of the Committee to Select Hour Speakers for Eastern Sectional Meetings; it was originally submitted to the Bulletin of the American Mathematical Society on October 29, 1952, but it could not be published among the invited addresses in the Bulletin because of its length.

[1] Part of the research on which this paper is based was done in connection with a contract with the Office of Naval Research.

Introduction.

1. Differential fields and meaning of normality. The study of algebraic equations has led to the concept of field, and thence to the beginnings of algebraic geometry and to Galois theory. In much the same way, the study of algebraic differential equations has, in modern times, led to the concept of *differential field*, and thence, in the work of the late J. F. Ritt, to the extensive theory of differential algebra, which in its elementary parts bears considerable analogy to the elementary parts of algebraic geometry (see Ritt [8]). A differential field is a commutative field, in the usual sense, together with a finite family of operators $\delta_1, \cdots, \delta_m$ each of which maps the field into itself as a derivation and which commute in pairs; a differential field is said to be *ordinary* or *partial* according as the number m equals or exceeds 1.[2] *In the present paper the expression "differential field" always stands for "differential field of characteristic* 0." The purpose of the present paper is to develop a Galois theory for such differential fields.

A main problem in initiating such a theory is to find a suitable definition of normal extension of a differential field. Now, two special cases of a Galois theory already exist, and it is natural to look to these examples for hints, and to require that any general theory developed generalize these two cases. One of these cases is the Galois theory of differential field extensions of finite degree, which is the classical Galois theory, since a relative field isomorphism of such an extension is automatically a differential field isomorphism. The other case is the Picard-Vessiot theory (see Kolchin [3] and [6]; a certain

[2] By permitting m to be 0 it is possible to subsume the concept of field under that of differential field; we shall not pursue this possibility further here, and when we refer to a differential field it will always be understood that $m \geq 1$.

familiarity of the reader with the contents of these two papers will be assumed).

In the classical Galois theory an algebraic field extension of characteristic 0 is normal if the field of invariants of the group of all automorphisms of the extension over the ground-field is the ground-field itself. If an extension has this property it follows that it has the same property when considered as an extension of any intermediate field; indeed, the fundamental theorem of Galois theory could not hold were this not the case. When we turn to differential fields, however, the state of affairs is different. If $\mathfrak{F}$ is a differential field and $\mathfrak{G}$ is an extension of $\mathfrak{F}$ such that every invariant of the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ belongs to $\mathfrak{F}$, and if $\mathfrak{F}_1$ is a differential field between $\mathfrak{F}$ and $\mathfrak{G}$, it does not follow, even if $\mathfrak{G}$ is finitely generated and differentially algebraic over $\mathfrak{F}$, that every invariant of the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}_1$ belongs to $\mathfrak{F}_1$ (see the example in footnote 7). Accordingly, we define $\mathfrak{G}$ to be *weakly normal* over $\mathfrak{F}$ if the invariants of the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ all belong to $\mathfrak{F}$, and $\mathfrak{G}$ to be *normal* over $\mathfrak{F}$ if $\mathfrak{G}$ is weakly normal over every differential field between $\mathfrak{F}$ and $\mathfrak{G}$. The latter is the same definition as given in Kolchin [3], § 16; in that paper it was shown, and indeed it is obvious, that when $\mathfrak{G}$ is normal over $\mathfrak{F}$ in this sense then there is a one-to-one Galois correspondence between the set of all differential fields intermediate to $\mathfrak{F}$ and $\mathfrak{G}$ and a *certain* set of subgroups of the group $\mathfrak{G}$ of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$. If the subgroup corresponding to an intermediate differential field is normal then the intermediate differential field is a normal extension of $\mathfrak{F}$ (but not conversely!). Aside from the fact that in this definition of normality we demand what is essentially the conclusion of the theorem we wish to prove, there remain two blemishes, one of which we can remove, the other of which we can not. The first blemish is that, when the subgroup $\mathfrak{G}(\mathfrak{F}_1)$ of $\mathfrak{G}$ corresponding to $\mathfrak{F}_1$ is normal, so that $\mathfrak{F}_1$ is a normal extension of $\mathfrak{F}$, the factor group $\mathfrak{G}/\mathfrak{G}(\mathfrak{F}_1)$ need not be isomorphic with the group of all automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$. This situation is remedied by defining a set of isomorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ to be *abundant* if, for every intermediate differential field $\mathfrak{F}_1$ and every element α of $\mathfrak{G}$ not in $\mathfrak{F}_1$, there exists an isomorphism σ in the set which leaves every element of $\mathfrak{F}_1$ invariant but which does not leave α invariant; clearly $\mathfrak{G}$ is normal if and only if the group of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ is abundant. Furthermore, if $\mathfrak{G}$ is normal over $\mathfrak{F}$, and $\mathfrak{G}$ is *any* abundant group of automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ (not necessarily the full automorphism group), the above mentioned results continue to hold and, when $\mathfrak{G}(\mathfrak{F}_1)$ is a normal subgroup of $\mathfrak{G}$, so that $\mathfrak{F}_1$ is a normal extension of $\mathfrak{F}$, then $\mathfrak{G}/\mathfrak{G}(\mathfrak{F}_1)$ is isomorphic to an abundant group of automorphisms of $\mathfrak{F}_1$ over $\mathfrak{F}$. The

second and more serious blemish is that we have no characterization of those "certain" subgroups which correspond to the intermediate differential fields. To avoid this defect we seek a more stringent definition.

In the classical Galois theory a normal extension is characterized also by the property that every relative isomorphism of the extension into any overfield of the extension is actually an automorphism. It would be unreasonable to demand the analogous property for differential fields, as this would exclude even the Picard-Vessiot extensions; indeed it can be shown that the extension would then be a normal algebraic extension in the classical sense. However, a hint of how to proceed is contained in the Picard-Vessiot theory. Let $\mathcal{G}$ be a Picard-Vessiot extension of $\mathcal{F}$, presupposing thereby that $\mathcal{F}$ and $\mathcal{G}$ are subject to the restrictions that $\mathcal{F}$ and $\mathcal{G}$ have the same field of constants $\mathcal{C}$, that $\mathcal{C}$ is algebraically closed, and that $\mathcal{G}$ is finitely generated and of finite transcendence degree over $\mathcal{F}$; it is easy to verify that $\mathcal{F}$ and $\mathcal{G}$ have the property that if σ is any isomorphism of $\mathcal{G}$ over $\mathcal{F}$ into an extension of $\mathcal{G}$ and if $\mathcal{C}_\sigma$ denotes the field of constants of the compositum $\mathcal{G}\langle\sigma\mathcal{G}\rangle$ then

$$(1) \qquad \mathcal{G}\langle\sigma\mathcal{G}\rangle = \mathcal{G}\langle\mathcal{C}_\sigma\rangle = (\sigma\mathcal{G})\langle\mathcal{C}_\sigma\rangle.$$

This is the property which we use, in the general case subject to the above restrictions, for our definition, which may be formulated in the following manner. We define an isomorphism σ of $\mathcal{G}$ into an extension of $\mathcal{G}$ to be *strong* if (1) holds; obviously every automorphism is a strong isomorphism. We then say, when $\mathcal{F}$ and $\mathcal{G}$ are subject to the above restrictions, that $\mathcal{G}$ is *strongly normal* over $\mathcal{F}$ if every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is strong.[3] As indicated in the following summary, it is this type of normality which appears to be the fruitful one.

2. Summary. Chapter I contains various results from elementary differential algebra which are used in the succeeding chapters. Several of these extend to partial differential fields results which are already known in the ordinary case. One theorem proved asserts the existence, for any differential field, of a suitably defined *universal extension*; roughly speaking, a given extension is universal if it is so big that all elements of all extensions we ever have occasion to introduce may be taken in the given extension. The use of a universal extension, which follows the now well-known procedure of

---

[3] By pursuing further the possibility mentioned in footnote [2] and suitably formulating the definition of constant, we could make the classical concept of normal algebraic extension of finite degree of a field of characteristic 0 a special case of concept of strongly normal extension of a differential field.

modern algebraic geometry (see Weil [9]), makes it possible to avoid certain logical difficulties connected with phrases like "the set of all extensions" (of a given differential field).

Chapter II contains a detailed study of strong isomorphisms of an extension $\mathcal{G}$ of a differential field $\mathcal{F}$, subject to the restrictions above. It is shown that it is possible, in a natural way, to introduce a multiplication in the set $\mathfrak{G}^*$ of all strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, with respect to which $\mathfrak{G}^*$ becomes a group; the group $\mathfrak{G}$ of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is then a subgroup of $\mathfrak{G}^*$. The concept of specialization is defined for an isomorphims of $\mathcal{G}$ into an extension of $\mathcal{G}$ (or more generally, for a family of such isomorphisms): σ' is called a *specialization* of σ if the family of elements $(\sigma'\alpha)_{\alpha \varepsilon G}$ is a specialization over $\mathcal{G}$ of the family $(\sigma\alpha)_{\alpha\varepsilon G}$. If σ is a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$ so is every specialization of σ. Most of the important facts concerning specializations of strong isomorphisms follow from Proposition 9 of Chapter II, which asserts that if $\sigma_1, \cdots, \sigma_p$ are strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ and if $\gamma_{ik} \varepsilon \mathcal{E}_{\sigma_i} (1 \leq i \leq p, 1 \leq k \leq q_i)$ then, roughly speaking, a specialization of $(\gamma_{ik})_{1 \leq i \leq p, 1 \leq k \leq q_i}$ over $\mathcal{E}$ can, under certain general conditions, be extended to a specialization of $(\sigma_1, \cdots, \sigma_p)$ in such a way that various inequalities are preserved. An algebraico-geometric structure is introduced into $\mathfrak{G}^*$ in the following way. A subset $\mathfrak{M}^*$ of $\mathfrak{G}^*$ is called an *irreducible set* in $\mathfrak{G}^*$ if $\mathfrak{M}^*$ contains an element σ^* such that $\mathfrak{M}^*$ is the set of all specializations of σ^*; σ^* is then called a *generic element* of $\mathfrak{M}^*$, and the transcendence degree of $\mathcal{G}\langle\sigma^*\mathcal{G}\rangle$ over $\mathcal{G}$, which is the same as the transcendence degree of $\mathcal{E}_{\sigma^*}$ over $\mathcal{E}$ and does not depend on the choice of generic element σ^*, is called the *dimension* of $\mathfrak{M}^*$. A subset $\mathfrak{M}^*$ of $\mathfrak{G}^*$ is called an *algebraic set* in $\mathfrak{G}^*$ if $\mathfrak{M}^*$ is the union of a finite set of irreducible sets in $\mathfrak{G}^*$; the definition and elementary properties of the components of an algebraic set quickly follow. This algebraico-geometric structure in $\mathfrak{G}^*$ induces a similar structure in $\mathfrak{G}$. Some propositions are proved about algebraic sets in $\mathfrak{G}$ which are analogous to some elementary results in algebraic geometry. Finally, by combining the group structure and algebraico-geometric structure of $\mathfrak{G}$ we arrive at the concept of *algebraic group* in $\mathfrak{G}$. Several simple results about such algebraic groups are proved which are like certain known results on algebraic matric groups (Kolchin [3]) and, more generally, group varities in the sense of Weil [10].

In Chapter III, after a brief discussion of normal extensions of differential fields, the results of Chapter II are applied to develop a Galois theory of strongly normal extensions. It is shown that strong normality implies normality, but not conversely. Let $\mathcal{G}$ be a strongly normal extension of $\mathcal{F}$.

The group $\mathfrak{G}$ of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is itself algebraic, and there is a one-to-one Galois correspondence between the set of all intermediate differential fields and a certain set of subgroups of $\mathfrak{G}$; this certain set is characterized as the set of all algebraic groups in $\mathfrak{G}$. The transcendence degree of $\mathcal{G}$ over any intermediate differential field $\mathcal{F}_1$ is proved to be equal to the dimension of the corresponding group $\mathfrak{G}(\mathcal{F}_1)$; the component of $\mathfrak{G}$ containing the identity (which component is unique and is a normal algebraic subgroup of $\mathfrak{G}$ of finite index) corresponds to the relative algebraic closure of $\mathcal{F}$ in $\mathcal{G}$. It is shown that if $\mathcal{F}_1$ is an intermediate differential field then the following conditions are equivalent: 1) $\mathcal{F}_1$ is strongly normal over $\mathcal{F}$; 2) $\mathcal{F}_1$ is normal over $\mathcal{F}$; 3) $\mathcal{F}_1$ is weakly normal over $\mathcal{F}$; 4) $\sigma\mathcal{F}_1 \subseteq \mathcal{F}_1$ for every $\sigma \varepsilon \mathfrak{G}$; 5) $\mathfrak{G}(\mathcal{F}_1)$ is a normal subgroup of $\mathfrak{G}$. And when these conditions are satisfied, the factor group $\mathfrak{G}/\mathfrak{G}(\mathcal{F}_1)$ is isomorphic with the group of all automorphisms of $\mathcal{F}_1$ over $\mathcal{F}$.

The remainder of Chapter III is devoted to three special types of extension. An element α is defined to be *primitive* over a differential field $\mathcal{F}$ if $\delta_i\alpha \varepsilon \mathcal{F}$ $(1 \leq i \leq m)$, to be *exponential* over $\mathcal{F}$ if $\alpha \neq 0$ and $\alpha^{-1}\delta_i\alpha \varepsilon \mathcal{F}$ $(1 \leq i \leq m)$, and to be *weierstrassian* over $\mathcal{F}$ if α is not a constant, and there exist two elements $g_2, g_3 \varepsilon \mathcal{C}$ with the polynomial $4y^3 - g_2y - g_3$ having simple roots only and m elements $a_1, \cdots, a_m \varepsilon \mathcal{F}$ such that $(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3)$ $(1 \leq i \leq m)$. In all three cases, if the field of constants of $\mathcal{F}\langle\alpha\rangle$ is $\mathcal{C}$, $\mathcal{F}\langle\alpha\rangle$ is strongly normal over $\mathcal{F}$. In the first two cases $\mathcal{F}\langle\alpha\rangle$ is a Picard-Vessiot extension of $\mathcal{F}$, but in the third case it is not, unless it is algebraic; indeed, it can be shown that if α is weierstrassian over $\mathcal{F}$ and if it is possible to find a family $\mathcal{F}_0, \mathcal{F}_1, \cdots, \mathcal{F}_r$ of differential fields such that $\mathcal{F}_0 = \mathcal{F}$, $\mathcal{F}_i$ is a Picard-Vessiot extension of $\mathcal{F}_{i-1}$ $(1 \leq i \leq r)$, and $\alpha \varepsilon \mathcal{F}_r$, then α is algebraic over $\mathcal{F}$. Reciprocally, it is shown that if a Picard-Vessiot extension can be obtained by a sequence of adjunctions of algebraic, primitive, exponential, and weierstrassian elements, then it can be obtained by adjunction of algebraic, primitive, and exponential elements alone. When α is transcendental over $\mathcal{F}$ then, in all three cases, $\mathcal{F}\langle\alpha\rangle$ is of transcendence degree 1 over $\mathcal{F}$. Conversely, it is proved that every strongly normal (and indeed every weakly normal) extension of $\mathcal{F}$ of transcendence degree 1 can be obtained from $\mathcal{F}$ by combining with algebraic adjunctions an adjunction of one of these three types. The proof of this converse, which is long and in places involves complicated computations, makes use of the well-known theorem that the group of automorphisms of an algebraic function field of one variable over an algebraically closed field of characteristic 0 is finite if the genus exceeds 1.

3. Problems. Various problems remain for investigation; we mention three, which are related.

First, there is the connection between algebraic groups of automorphisms as defined herein, and group varieties as defined by A. Weil. Is it always possible to identify the component of the identity of an algebraic group with a group variety? Conversely, is every group variety identifiable with an algebraic group?

Second, there is the task of characterizing, if possible, by algebraic-group properties, those strongly normal extensions which are Picard-Vessiot extensions.

Third, there is the significance of solvability, or even of commutativity, of the group of automorphisms of a strongly normal extension. For what sort of strongly normal extension is the group abelian? It is conceivable that investigation of this question will lead into the theory of abelian functions.

4. Notation. The notation used is more or less the same as in Kolchin [3], and is reasonably standard. We mention only that the degree of transcendence and the degree of differential transcendence of $\mathcal{G}$ over $\mathcal{F}$ are denoted by $\partial^0\mathcal{G}/\mathcal{F}$ and $\nabla^0\mathcal{G}/\mathcal{F}$ respectively.

Chapter I. Differential-algebraic preliminaries.

1. A lemma on polynomial ideals. Let K be a field of characteristic 0,[4] and let $y_1, \cdots, y_n$ be indeterminates. We shall prove the following lemma, which collects certain known facts in a form convenient for future use.

LEMMA. *Let $\mathfrak{p}$ be a prime ideal of $K[y_1, \cdots, y_n]$ of dimension d. For every extension L of K the ideal $L \cdot \mathfrak{p}$ generated by $\mathfrak{p}$ in $L[y_1, \cdots, y_n]$ is equal to its own radical; the minimal prime ideal divisors $\mathfrak{p}_1, \cdots, \mathfrak{p}_r$ of $L \cdot \mathfrak{p}$ all have dimension d; every generic zero of every $\mathfrak{p}_i$ is a generic zero of $\mathfrak{p}$, and every generic zero of $\mathfrak{p}$ is a zero of precisely one $\mathfrak{p}_i$. There exists, independent of L, an irreducible polynomial R with coefficients in K such that for every extension L of K the number of minimal prime ideal divisors of $L \cdot \mathfrak{p}$ equals the number of irreducible factors into which R splits over L.*

Proof. Let $\mathfrak{P}$ be the radical of $L \cdot \mathfrak{p}$, so that $\mathfrak{P} = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_r$ where $\mathfrak{p}_1, \cdots, \mathfrak{p}_r$ are the minimal prime ideal divisors of $L \cdot \mathfrak{p}$. Let $(\eta_{i1}, \cdots, \eta_{in})$ be a generic zero of $\mathfrak{p}_i$; $(\eta_{i1}, \cdots, \eta_{in})$ is obviously a zero of $\mathfrak{p}$. If

[4] This condition, which suffices for the present purposes, can be relaxed.

$F \in K[y_1, \cdots, y_n]$ vanishes at $(\eta_{i1}, \cdots, \eta_{in})$ then $F \in \mathfrak{p}_i$, so that if we let G be a polynomial in $L[y_1, \cdots, y_n]$ such that $G \in \bigcap_{j \neq i} \mathfrak{p}_j$, $G \notin \mathfrak{p}_i$ then $FG \in \mathfrak{P}$, whence for some exponent $e > 0$ we have $F^e G^e \in L \cdot \mathfrak{p}$. Therefore there exist elements $\lambda_k \in L$ linearly independent over K such that we may write $F^e G^e = \Sigma \lambda_k P_k$, where each $P_k \in \mathfrak{p}$, and $G^e = \Sigma \lambda_k G_k$, where each $G_k \in K[y_1, \cdots, y_n]$. From this we see that $\Sigma \lambda_k F^e G_k = \Sigma \lambda_k P_k$, so that $F^e G_k = P_k \in \mathfrak{p}$ for every k; since not every G_k belongs to $\mathfrak{p}$ (for otherwise G would belong to $\mathfrak{p}_i$) and since $\mathfrak{p}$ is prime we conclude that $F \in \mathfrak{p}$. This shows that $(\eta_{i1}, \cdots, \eta_{in})$ is a generic zero of $\mathfrak{p}$.

Let $(\eta_1, \cdots, \eta_n)$ be any generic zero of $\mathfrak{p}$. For the sake of definiteness we suppose that $\eta_1, \cdots, \eta_d$ are algebraically independent over K; then there exists an element ω such that $K(\eta_1, \cdots, \eta_n) = K(\eta_1, \cdots, \eta_d, \omega)$. Let w be a new indeterminate and let R be a polynomial in $K[y_1, \cdots, y_d, w]$ of as low degree as possible which vanishes at $(\eta_1, \cdots, \eta_d, \omega)$, so that R is irreducible over K. Because $(\eta_{i1}, \cdots, \eta_{in})$ is also a generic zero of $\mathfrak{p}$, $(\eta_{i1}, \cdots, \eta_{in})$ is a generic specialization [5] of $(\eta_1, \cdots, \eta_n)$ over K, and therefore can be extended to a generic specialization $(\eta_{i1}, \cdots, \eta_{in}, \omega_i)$ of $(\eta_1, \cdots, \eta_n, \omega)$ over K. Now $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a zero of R and therefore of some irreducible factor of R over L; moreover $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a zero of only one of these irreducible factors, for otherwise $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ and therefore $(\eta_1, \cdots, \eta_d, \omega)$ would be a zero of $\partial R/\partial w$ which is of lower degree than R. We denote the irreducible factor of R over L which vanishes at $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ by R_i.

Let $(\eta'_{i1}, \cdots, \eta'_{id}, \omega'_i)$ be a generic zero of the prime ideal $\mathfrak{r}_i$ generated by R_i in $L[y_1, \cdots, y_d, w]$. Then $(\eta_{i1}, \cdots, \eta_{id}, \omega_i)$ is a specialization of $(\eta'_{i1}, \cdots, \eta'_{id}, \omega'_i)$ over L and a generic specialization of $(\eta'_{i1}, \cdots, \eta'_{id}, \omega'_i)$ over K; therefore there exist elements $\eta'_{i,d+1}, \cdots, \eta'_{in}$ such that $(\eta_{i1}, \cdots, \eta_{in}, \omega_i)$ is a generic specialization of $(\eta'_{i1}, \cdots, \eta'_{in}, \omega'_i)$ over K. Now $(\eta'_{i1}, \cdots, \eta'_{in})$ is a zero of $\mathfrak{p}$ and therefore of $\mathfrak{p}_j$ for some j; since $(\eta_{i1}, \cdots, \eta_{in})$ must be a zero of this $\mathfrak{p}_j$ and since $\mathfrak{p}_j \not\subset \mathfrak{p}_i$ if $i \neq j$ it follows that $i = j$. As $\eta'_{i1}, \cdots, \eta'_{id}$ are obviously algebraically independent over L we have

$$\partial^0 L(\eta'_{i1}, \cdots, \eta'_{in})/L \geq d = \partial^0 K(\eta_{i1}, \cdots, \eta_{in})/K$$
$$\geq \partial^0 L(\eta_{i1}, \cdots, \eta_{in})/L = \dim \mathfrak{p}_i,$$

so that $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic zero of $\mathfrak{p}_i$ and $\dim \mathfrak{p}_i = d$. It follows that $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic specialization of $(\eta_{i1}, \cdots, \eta_{in})$ over L, so that

[5] The very elementary facts concerning specializations over a field used in this paper can be found in Weil [9], chapter II.

$(\eta'_{i1},\cdots,\eta'_{in},\omega'_i)$ is a generic specialization of $(\eta_{i1},\cdots,\eta_{in},\omega_i)$ over L, and $(\eta_{i1},\cdots,\eta_{id},\omega_i)$ is a generic zero of $\mathfrak{r}_i$.

If $R_i=R_j$ then $\mathfrak{r}_i=\mathfrak{r}_j$ and $(\eta_{i1},\cdots,\eta_{id},\omega_i)$ is a generic specialization of $(\eta_{j1},\cdots,\eta_{jd},\omega_j)$ over L, so that $(\eta_{i1},\cdots,\eta_{in})$ is a generic specialization of $(\eta_{j1},\cdots,\eta_{jn})$ over L, whence $\mathfrak{p}_i=\mathfrak{p}_j$ and $i=j$. Thus $R_1,\cdots,R_r$ are distinct irreducible factors of R over L. Let S be any irreducible factor of R over L, let $\mathfrak{s}$ denote the prime ideal generated by S in $L[y_1,\cdots,y_d,w]$, and let $(\zeta_1,\cdots,\zeta_d,\theta)$ be a generic zero of $\mathfrak{s}$; $(\zeta_1,\cdots,\zeta_d,\theta)$ is clearly a generic specialization of $(\eta_1,\cdots,\eta_d,\omega)$ over K and therefore can be extended to a generic specialization $(\zeta_1,\cdots,\zeta_n,\theta)$ of $(\eta_1,\cdots,\eta_n,\omega)$. $(\zeta_1,\cdots,\zeta_n)$ is a zero of $\mathfrak{p}$ and therefore of $\mathfrak{p}_i$ for some i; therefore $(\zeta_1,\cdots,\zeta_d,\theta)$ is a zero of R_i, so that R_i is divisible by S. It follows that $R_1,\cdots,R_r$ are all the irreducible factors of R over L, so that the number of minimal prime ideal divisors of $L\cdot\mathfrak{p}$ equals the number of irreducible factors of R over L. Also $(\eta_1,\cdots,\eta_n)$, which obviously is a zero of some $\mathfrak{p}_i$, is a zero of only one $\mathfrak{p}_i$; for if $(\eta_1,\cdots,\eta_n)$ were a zero of $\mathfrak{p}_i$ and $\mathfrak{p}_j$ $(i\neq j)$ then $(\eta_1,\cdots,\eta_d,\omega)$ would be a zero of R_i and R_j, and therefore of $\partial R/\partial w$, which is of lower degree than R.

It remains to prove that $\mathfrak{P}=L\cdot\mathfrak{p}$, and to do this it suffices to show that $\mathfrak{P}\subseteq L\cdot\mathfrak{p}$. If $F\in\mathfrak{P}$ then we may write $F=\Sigma\lambda_jF_j$, where each F_j belongs to $K[y_1,\cdots,y_n]$ and (λ_j) is a family of elements of L linearly independent over K. Let $A_{d+1},B_{d+1},\cdots,A_n,B_n$ be polynomials in $K[y_1,\cdots,y_d,w]$ such that

$$\eta_k=A_k(\eta_1,\cdots,\eta_d,\omega)/B_k(\eta_1,\cdots,\eta_d,\omega),\ d+1\leq k\leq n.$$

Then there exists a single exponent $e\geq 0$ such that for each j

$$(B_{d+1}\cdots B_n)^eF_j\equiv G_j(B_{d+1}y_{d+1}-A_{d+1},\cdots,B_ny_n-A_n),$$

where $G_j\in K[y_1,\cdots,y_d,w]$. It is easy to see that $\Sigma\lambda_jG_j$ vanishes at $(\eta_{i1},\cdots,\eta_{id},\omega_i)$ for each i and therefore is divisible by R. Because the λ_j's are linearly independent over K it easily follows that each G_j is divisible by R, so that each G_j vanishes at $(\eta_1,\cdots,\eta_d,\omega)$, each $(B_{d+1}\cdots B_n)^eF_j$ vanishes at $(\eta_1,\cdots,\eta_n,\omega)$, each F_j vanishes at $(\eta_1,\cdots,\eta_n)$, each $F_j\in\mathfrak{p}$, and $F\in L\cdot\mathfrak{p}$.

2. Prime differential ideals and differential field extension. Let $\mathcal{F}$ be a differential field and let $y_1,\cdots,y_n$ denote indeterminates. If Π is a prime differential ideal of the differential ring $\mathcal{F}\{y_1,\cdots,y_n\}$ and $(\eta_1,\cdots,\eta_n)$ is a generic zero of Π then the degree of differential transcendence

$\nabla^0\mathcal{F}\langle\eta_1,\cdots,\eta_n\rangle/\mathcal{F}$ is called the *dimension* of Π (notation: $\dim\Pi$); $\dim\Pi$ does not depend on the particular generic zero used, and $0\leq\dim\Pi\leq n$. The degree of transcendence $\partial^0\mathcal{F}\langle\eta_1,\cdots,\eta_n\rangle/\mathcal{F}$ is called the *order* of Π (notation: $\operatorname{ord}\Pi$); $\operatorname{ord}\Pi$, which does not depend on the particular generic zero used, either is an integer ≥ 0 (finite order) or is ∞ (infinite order). The following result is well-known for the case in which $\mathcal{F}$ is an ordinary differential field (see e. g. Ritt [8], pp. 50-51).

PROPOSITION 1. *Let Π be a prime differential ideal of $\mathcal{F}\{y_1,\cdots,y_n\}$. For every extension $\mathcal{G}$ of $\mathcal{F}$ the ideal $\mathcal{G}\cdot\Pi$ generated by Π in $\mathcal{G}\{y_1,\cdots,y_n\}$ is a perfect differential ideal; the minimal prime differential ideal divisors $\Pi_1,\cdots,\Pi_r$ of $\mathcal{G}\cdot\Pi$ all have the same dimension as Π, and all have the same order as Π; every generic zero of every Π_i is a generic zero of Π, and every generic zero of Π is a zero of precisely one Π_i. There exists, independent of $\mathcal{G}$, an irreducible polynomial R with coefficients in $\mathcal{F}$ such that for every extension $\mathcal{G}$ of $\mathcal{F}$ the number of minimal prime differential ideal divisors of $\mathcal{G}\cdot\Pi$ equals the number of irreducible factors into which R splits over $\mathcal{G}$.*

Proof. Let $\mathcal{R}$, $\mathcal{R}'$ denote $\mathcal{F}\{y_1,\cdots,y_n\}$, $\mathcal{G}\{y_1,\cdots,y_n\}$ respectively. For each integer $k\geq 0$ let $\mathcal{R}_k$, $\mathcal{R}'_k$ denote the set of all elements of $\mathcal{R}$, $\mathcal{R}'$ respectively which do not have order $>k$. We shall consider $\mathcal{R}_k$ and $\mathcal{R}'_k$ as polynomial rings over $\mathcal{F}$ and $\mathcal{G}$ respectively (each element of $\mathcal{R}_k$ and of $\mathcal{R}'_k$ is a polynomial in the expressions $\delta_1^{i_1}\cdots\delta_m^{i_m}y_j$ with $0\leq i_1+\cdots+i_m\leq k$, $1\leq j\leq n$).

For every $k\geq 0$, $\Pi\cap\mathcal{R}_k$ is a prime ideal of $\mathcal{R}_k$, so that (§ 1) $\mathcal{G}\cdot(\Pi\cap\mathcal{R}_k)$ is an ideal of $\mathcal{R}'_k$ which is equal to its own radical. Clearly $\mathcal{G}\cdot\Pi$ is a differential ideal of $\mathcal{R}'$; if $\mathcal{G}\cdot\Pi$ were not perfect there would exist an $F\in\mathcal{R}'$ and an integer $e>0$ with $F\notin\mathcal{G}\cdot\Pi$, $F^e\in\mathcal{G}\cdot\Pi$, so that for large k we would have $F\notin\mathcal{G}\cdot(\Pi\cap\mathcal{R}_k)$, $F^e\in\mathcal{G}\cdot(\Pi\cap\mathcal{R}_k)$, contradicting the fact that $\mathcal{G}\cdot(\Pi\cap\mathcal{R}_k)$ is its own radical. Thus $\mathcal{G}\cdot\Pi$ is a perfect differential ideal of $\mathcal{R}'$.

We now assert that $\mathcal{G}\cdot(\Pi\cap\mathcal{R}_k)=(\mathcal{G}\cdot\Pi)\cap\mathcal{R}'_k$. Indeed, it is obvious that $\mathcal{G}\cdot(\Pi\cap\mathcal{R}_k)\subseteq(\mathcal{G}\cdot\Pi)\cap\mathcal{R}'_k$. Suppose then that $F\in(\mathcal{G}\cdot\Pi)\cap\mathcal{R}'_k$. Then we may write

$$F=\sum P_k\phi_k \tag{1}$$

where the elements $\phi_k\in\mathcal{G}$ are linearly independent over $\mathcal{F}$ and each $P_k\in\Pi$. Fixing our attention on any derivative $\delta_1^{i_1}\cdots\delta_m^{i_m}y_j$ of order $i_1+\cdots+i_m>k$, let C_k denote the coefficient in P_k of any fixed positive power of this derivative; since F is not of order $>k$, (1) yields the relation $\sum C_k\phi_k=0$, so that each

$C_k = 0$. It follows that no P_k has order $> k$, so that each $P_k \varepsilon \mathcal{R}_k$, and $F \varepsilon \mathcal{G} \cdot (\Pi \cap \mathcal{R}_k)$. This proves our assertion.

Since the perfect differential ideal $\mathcal{G} \cdot \Pi$ is the intersection of its minimal prime differential ideal divisors $\Pi_1, \cdots, \Pi_r$ it is a consequence of the above assertion that

$$\mathcal{G} \cdot (\Pi \cap \mathcal{R}_k) = (\Pi_1 \cap \mathcal{R}'_k) \cap \cdots \cap (\Pi_r \cap \mathcal{R}'_k).$$

Now $\Pi_i \not\subset \Pi_j$ if $i \neq j$, so that if k is sufficiently great $\Pi_i \cap \mathcal{R}'_k \not\subset \Pi_j \cap \mathcal{R}'_k$ whenever $i \neq j$. Taking k large enough for this to be the case, we see that $\Pi_1 \cap \mathcal{R}'_k, \cdots, \Pi_r \cap \mathcal{R}'_k$ are the minimal prime ideal divisors of $\mathcal{G} \cdot (\Pi \cap \mathcal{R}_k)$ in $\mathcal{R}'_k$.

It is obvious that Π_i is of dimension no higher than Π. If Π_i had lower dimension than Π then there would exist a subset $z_1, \cdots, z_t$ of $y_1, \cdots, y_n$ such that Π contains no nonzero differential polynomial in $z_1, \cdots, z_t$ alone but Π_i does, that is (for k large) $\Pi \cap \mathcal{R}_k$ contains no nonzero polynomial in the expressions $\delta_1^{i_1} \cdots \delta_m^{i_m} z_l$ with $0 \leq i_1 + \cdots + i_m \leq k$, $1 \leq l \leq t$, but $\Pi_i \cap \mathcal{R}'_k$ does; this would imply that the prime polynomial ideal $\Pi_i \cap \mathcal{R}'_k$ of $\mathcal{R}'_k$ has lower dimension than the prime polynomial ideal $\Pi \cap \mathcal{R}_k$ of $\mathcal{R}_k$, contradicting the lemma of § 1. Therefore each Π_i is of the same dimension as Π.

Again, the order of Π_i obviously equals the limit (including the possibility ∞) as k becomes infinite of the dimension of $\Pi_i \cap \mathcal{R}'_k$, which by the lemma equals the limit of the dimension of $\Pi \cap \mathcal{R}_k$, which equals the order of Π.

If $(\eta_{i1}, \cdots, \eta_{in})$ is a generic zero of Π_i then obviously

$$(\delta_1^{i_1} \cdots \delta_m^{i_m} \eta_{ij})_{0 \leq i_1 + \ldots + i_m \leq k,\ 1 \leq j \leq n}$$

is a generic zero of $\Pi_i \cap \mathcal{R}'_k$ and therefore (by the lemma) a generic zero of $\Pi \cap \mathcal{R}_k$; since k can be arbitrarily large this means that $(\eta_{i1}, \cdots, \eta_{in})$ is a generic zero of Π.

If $(\eta_1, \cdots, \eta_n)$ is a generic zero of Π then $(\eta_1, \cdots, \eta_n)$ is a zero of at least one Π_i; if it were a zero of Π_i for two distinct values of i then for k large $(\delta_1^{i_1} \cdots \delta_m^{i_m} \eta_j)_{0 \leq i_1 + \ldots + i_m \leq k,\ 1 \leq j \leq n}$ would be a generic zero of $\Pi \cap \mathcal{R}_k$ and a zero of $\Pi_i \cap \mathcal{R}'_k$ for two distinct values of i, contradicting the lemma.

It remains to prove the existence of a polynomial R as described in the statement of the proposition. To this end let $k(\mathcal{G})$ be the smallest integer such that, for all $k \geq k(\mathcal{G})$, $\Pi_1 \cap \mathcal{R}'_k, \cdots, \Pi_r \cap \mathcal{R}'_k$ are the minimal prime ideal divisors of $\mathcal{G} \cdot (\Pi \cap \mathcal{R}_k)$, that is such that, for all $k \geq k(\mathcal{G})$, $\Pi_i \cap \mathcal{R}'_k \not\subset \Pi_j \cap \mathcal{R}'_k$ whenever $i \neq j$. We shall show below that $k(\mathcal{G})$ is

an increasing function of $\mathcal{G}$, that is, if $\mathcal{H}$ is an extension of $\mathcal{G}$ then $k(\mathcal{G}) \leq k(\mathcal{H})$. Assuming this result, let us see how we can complete the proof of the proposition. By the lemma for each $k \geq 0$ there exists, independent of $\mathcal{G}$, an irreducible polynomial R_k with coefficients in $\mathcal{F}$ such that the number of minimal prime ideal divisors of $\mathcal{G} \cdot (\Pi \cap \mathcal{R}_k)$ in $\mathcal{R}'_k$ equals the number of irreducible factors of R_k over $\mathcal{G}$. Now the number of irreducible factors of R_k over $\mathcal{G}$ equals the number of irreducible factors of R_k over the relative algebraic closure $\mathcal{F}^0$ of $\mathcal{F}$ in $\mathcal{G}$. Therefore

number of minimal prime differential ideal divisors of $\mathcal{G} \cdot \Pi =$
number of minimal prime ideal divisors of $\mathcal{G} \cdot (\Pi \cap \mathcal{R}_k)$ (all $k \geq k(\mathcal{G})$) $=$
number of irreducible factors of R_k over $\mathcal{G}$ (all $k \geq k(\mathcal{G})$) $=$
number of irreducible factors of R_k over $\mathcal{F}^0$ (all $k \geq k(\mathcal{G})$) $=$
number of irreducible factors of R_k over $\mathcal{F}^0$ (all $k \geq k(\mathcal{F}^0)$) $=$
number of irreducible factors of R_k over $\mathcal{G}$ (all $k \geq k(\mathcal{F}^0)$).

Thus if we let $\mathcal{F}'$ be an algebraic closure [6] of $\mathcal{F}$ and set $R = R_l$, where $l = k(\mathcal{F}')$, then $l \geq k(\mathcal{F}^0)$ and R does not depend on $\mathcal{G}$, and the number of minimal prime differential ideal divisors of $\mathcal{G} \cdot \Pi$ equals the number of irreducible factors of R over $\mathcal{G}$.

We now show that $k(\mathcal{G})$ is an increasing function of $\mathcal{G}$. Let $\mathcal{H}$ be an extension of $\mathcal{G}$, let $\mathcal{R}'' = \mathcal{H}\{y_1, \cdots, y_n\}$, and let $\mathcal{R}''_k$ denote the ring of all elements of $\mathcal{R}''$ which do not have order $> k$. $\mathcal{H} \cdot \Pi$ is a perfect differential ideal of $\mathcal{R}''$ and $\mathcal{H} \cdot \Pi = \mathcal{H} \cdot (\mathcal{G} \cdot \Pi) = \mathcal{H} \cdot (\Pi_1 \cap \cdots \cap \Pi_r)$. Now if Π' is any ideal of R' and if for some $P'' \in \mathcal{H} \cdot \Pi'$ we write $P'' = \sum \phi_l P'_l$, where each $P'_l \in \mathcal{R}'$ and the elements ϕ_l of $\mathcal{H}$ are linearly independent over $\mathcal{G}$, then it is easy to see that each $P'_l \in \Pi'$. It follows that

$$(\mathcal{H} \cdot \Pi_1) \cap \cdots \cap (\mathcal{H} \cdot \Pi_r) = \mathcal{H} \cdot (\Pi_1 \cap \cdots \cap \Pi_r),$$

so that $\mathcal{H} \cdot \Pi = (\mathcal{H} \cdot \Pi_1) \cap \cdots \cap (\mathcal{H} \cdot \Pi_r)$. Now if $i_1 \neq i_2$ then a minimal prime differential ideal divisor Λ_1 of $\mathcal{H} \cdot \Pi_{i_1}$ can not be contained in a minimal prime differential ideal divisor Λ_2 of $\mathcal{H} \cdot \Pi_{i_2}$, because otherwise for every k we would have $\Lambda_2 \cap \mathcal{R}''_k \supseteq \mathcal{H} \cdot (\Pi_{i_1} \cap \mathcal{R}'_k + \Pi_{i_2} \cap \mathcal{R}'_k)$ and $\Lambda_2 \cap \mathcal{R}''_k$

[6] If K is any field of characteristic 0 and K' is an algebraic closure of K then every derivation of K has a unique extension which is a derivation of K' (see, for example, Bourbaki [1], chapter V, § 9, proposition 5, p. 139); moreover, it is easy to verify that if two derivations of K commute then there extended derivations of K' commute. It follows that every differential field $\mathcal{F}$ has an algebraically closed algebraic differential field extension; we call any such extension an *algebraic closure* of the differential field $\mathcal{F}$. Any two algebraic closures of $\mathcal{F}$ are isomorphic over $\mathcal{F}$.

would have lower dimension than $\Pi_{i_2} \cap \mathcal{R}'_k$ in contradiction to the lemma of § 1 (since for k large $\Lambda_2 \cap \mathcal{R}''_k$ is a minimal prime ideal divisor of $\mathcal{H} \cdot (\Pi_{i_2} \cap \mathcal{R}'_k)$). Therefore if we denote the minimal prime differential ideal divisors of $\mathcal{H} \cdot \Pi_i$ by $\Pi_{i1}, \cdots, \Pi_{i,s(i)}$ $(1 \leq i \leq r)$ then the minimal prime differential ideal divisors of $\mathcal{H} \cdot \Pi$ are the ideals Π_{ij} $(1 \leq i \leq r$, $1 \leq j \leq s(i))$. If $k < k(\mathcal{G})$ then there exist i, i' with $i \neq i'$ such that $\Pi_i \cap \mathcal{R}'_k \subseteq \Pi_{i'} \cap \mathcal{R}'_k$; for these i, i' we have

$$(\Pi_{i1} \cap \mathcal{R}''_k) \cap \cdots \cap (\Pi_{i,s(i)} \cap \mathcal{R}''_k) = (\mathcal{H} \cdot \Pi_i) \cap \mathcal{R}''_k$$
$$\subseteq (\mathcal{H} \cdot \Pi_{i'}) \cap \mathcal{R}''_k \subseteq \Pi_{i'1} \cap \mathcal{R}''_k,$$

so that for some j we have $\Pi_{ij} \cap \mathcal{R}''_k \subseteq \Pi_{i'1} \cap \mathcal{R}''_k$, whence $k < k(\mathcal{H})$. It follows that $k(\mathcal{G}) \leq k(\mathcal{H})$. As we have seen, this completes the proof of Proposition 1.

3. Specializations over differental fields. For purposes of convenience we extend the language of specializations, as used in algebraic geometry, to differential fields. Let $\mathcal{F}$ be a differential field and let $(\eta_j)_{j \varepsilon J}$ be an indexed family of elements of some extension of $\mathcal{F}$. A family $(\zeta_j)_{j \varepsilon J}$, with the same set of indices J, of elements of some extension of $\mathcal{F}$ will be called a *specialization of* $(\eta_j)_{j \varepsilon J}$ *over* $\mathcal{F}$ if, for every finite subset $j_1, \cdots, j_n$ of J, every differential polynomial in $\mathcal{F}\{y_1, \cdots, y_n\}$ which vanishes at $(\eta_{j_1}, \cdots, \eta_{j_n})$ also vanishes at $(\zeta_{j_1}, \cdots, \zeta_{j_n})$. If $(\zeta_j)_{j \varepsilon J}$ is a specialization of $(\eta_j)_{j \varepsilon J}$ over $\mathcal{F}$ such that $(\eta_j)_{j \varepsilon J}$ is a specialization of $(\zeta_j)_{j \varepsilon J}$ over $\mathcal{F}$ then we say that $(\zeta_j)_{j \varepsilon J}$ is a *generic* specialization of $(\eta_j)_{j \varepsilon J}$ over $\mathcal{F}$. If I is a subset of J and $(\zeta_j)_{j \varepsilon J}$ is a specialization of $(\eta_j)_{j \varepsilon J}$ over $\mathcal{F}$ then $(\zeta_i)_{i \varepsilon I}$ is a specialization of $(\eta_i)_{i \varepsilon I}$ over $\mathcal{F}$; we say in this case that the specialization $(\zeta_j)_{j \varepsilon J}$ of $(\eta_j)_{j \varepsilon J}$ over $\mathcal{F}$ is an *extension* of the specialization $(\zeta_i)_{i \varepsilon I}$ of $(\eta_i)_{i \varepsilon I}$ over $\mathcal{F}$. If $(\zeta_j)_{j \varepsilon J}$ is a generic specialization of $(\eta_j)_{j \varepsilon J}$ over $\mathcal{F}$ then there exists a unique isomorphism of $\mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle$ onto $\mathcal{F}\langle(\zeta_j)_{j \varepsilon J}\rangle$ over $\mathcal{F}$ which maps η_j onto ζ_j for every $j \varepsilon J$. If $(\zeta_j)_{j \varepsilon J}$ is a generic specialization of $(\eta_j)_{j \varepsilon J}$ over $\mathcal{F}$ and if $(\eta'_{j'})_{j' \varepsilon J'}$ is any family of elements of some extension of $\mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle$, then the specialization can be extended to a generic specialization

$$((\zeta_j)_{j \varepsilon J}, (\zeta'_{j'})_{j' \varepsilon J'}) \text{ of } ((\eta_j)_{j \varepsilon J}, (\eta'_{j'})_{j' \varepsilon J'})$$

over $\mathcal{F}$. The following proposition is well-known in the case of ordinary differential fields (Ritt [8], p. 49).

PROPOSITION 2. *If* $(\zeta_j)_{j \varepsilon J}$ *is a specialization of* $(\eta_j)_{j \varepsilon J}$ *over* $\mathcal{F}$ *then*

$$\nabla^0 \mathcal{F}\langle(\zeta_j)_{j \varepsilon J}\rangle/\mathcal{F} \leq \nabla^0 \mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle/\mathcal{F}$$

and

$$\partial^0 \mathcal{F}\langle(\zeta_j)_{j \varepsilon J}\rangle / \mathcal{F} \leq \partial^0 \mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle / \mathcal{F} ;$$

if in addition $\partial^0 \mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle / \mathcal{F}$ *is finite and equal to* $\partial^0 \mathcal{F}\langle(\zeta_j)_{j \varepsilon J}\rangle / \mathcal{F}$ *then the specialization is generic.*

Proof. The first part of the proposition is obvious. We prove the second part. Since $\partial^0 \mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle / \mathcal{F}$ and $\partial^0 \mathcal{F}\langle(\zeta_j)_{j \varepsilon J}\rangle / \mathcal{F}$ are finite and equal there exist a finite subset K of J such that

$$\partial^0 \mathcal{F}\langle(\zeta_j)_{j \varepsilon J'}\rangle / \mathcal{F} = \partial^0 \mathcal{F}\langle(\zeta_j)_{j \varepsilon J}\rangle / \mathcal{F} = \partial^0 \mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle / \mathcal{F} = \partial^0 \mathcal{F}\langle(\eta_j)_{j \varepsilon J'}\rangle / \mathcal{F}$$

for every finite subset J' of J which contains K. It is clear that $(\zeta_j)_{j \varepsilon J}$ is a generic specialization of $(\eta_j)_{j \varepsilon J}$ if $(\zeta_j)_{j \varepsilon J'}$ is a generic specialization of $(\eta_j)_{j \varepsilon J'}$ for every finite subset J' of J which contains K. It follows that we may assume that J is finite. Making this assumption, it is easy to see that there exists an integer $k_0 \geq 0$ such that

$$\mathcal{F}\langle(\eta_j)_{j \varepsilon J}\rangle = \mathcal{F}\left((\delta_1^{i_1} \cdots \delta_m^{i_m} \eta_j)_{0 \leq i_1 + \cdots + i_m \leq k,\, j \varepsilon J}\right),$$
$$\mathcal{F}\langle(\zeta_j)_{j \varepsilon J}\rangle = \mathcal{F}\left((\delta_1^{i_1} \cdots \delta_m^{i_m} \zeta_j)_{0 \leq i_1 + \cdots + i_m \leq k,\, j \varepsilon J}\right)$$

for every integer $k \geq k_0$. By a well-known result concerning specializations over a field (for example, see Weil [9], p. 28, Theorem 3) it follows that if we regard $\mathcal{F}$ as a field then $(\delta_1^{i_1} \cdots \delta_m^{i_m} \zeta_j)_{0 \leq i_1 + \cdots + i_m \leq k,\, j \varepsilon J}$ is a generic specialization of $(\delta_1^{i_1} \cdots \delta_m^{i_m} \eta_j)_{0 \leq i_1 + \cdots + i_m \leq k,\, j \varepsilon J}$ over $\mathcal{F}$ for every $k \geq k_0$. This implies that $(\zeta_j)_{j \varepsilon J}$ is a generic specialization of $(\eta_j)_{j \varepsilon J}$ over the differential field $\mathcal{F}$.

COROLLARY. *A zero* $(\zeta_1, \cdots, \zeta_n)$ *of a prime differential ideal* Π *of* $F\{y_1, \cdots, y_n\}$ *of finite order is generic if and only if*

$$\partial^0 \mathcal{F}\langle\zeta_1, \cdots, \zeta_n\rangle / \mathcal{F} = \operatorname{ord} \Pi .$$

4. Constants. Let $\mathcal{F}$ be a differential field, and denote the field of constants of $\mathcal{F}$ by $\mathcal{C}$. By the *order* of a differential operator $\delta_1^{i_1} \cdots \delta_m^{i_m}$ we mean the inteer $i_1 + \cdots + i_m$. If $\theta_1, \cdots, \theta_n$ are differential operators of the form $\delta_1^{i_1} \cdots \delta_m^{i_m}$ $(0 \leq i_1 < \infty, \cdots, 0 \leq i_m < \infty)$ then we use $W_{\theta_1 \ldots \theta_n}$ to denote the differential polynomial defined by $W_{\theta_1 \ldots \theta_n} = \det(\theta_i y_j)$.

PROPOSITION 3. *The elements* $\eta_1, \cdots, \eta_n$ *of* $\mathcal{F}$ *are linearly dependent over* $\mathcal{C}$ *if and only if* $W_{\theta_1 \ldots \theta_n}(\eta_1, \cdots, \eta_n) = 0$ *for all choices of* $\theta_1, \cdots, \theta_n$ *of order* $< n$.

This has been proved in Kolchin [6].

A consequence of this proposition is that if $\eta_1, \cdots, \eta_n$ are linearly dependent (or independent) over the field of constants of some differential field containing them then they are linearly dependent (or independent) over the field of constants of any differential field containing them; therefore we may speak simply of linear dependence or independence *over constants.*

COROLLARY 1. *Let L be a homogeneous linear polynomial in $\mathcal{F}[u_1, \cdots, u_q]$. There exist a finite number of homogeneous linear polynomials $L_1, \cdots, L_r$ in $\mathcal{C}[u_1, \cdots, u_q]$ such that q constants in an extension of $\mathcal{F}$ form a zero of L if and only if they form a zero of $L_1, \cdots, L_r$.*

Proof. Write $L = \sum_{i=1}^{r} L_i\alpha_i$, where $\alpha_1, \cdots, \alpha_r$ are elements of $\mathcal{F}$ linearly independent over $\mathcal{C}$ (and therefore over constants) and each L_i is a homogeneous linear polynomial in $\mathcal{C}[u_1, \cdots, u_q]$. If $\gamma_1, \cdots, \gamma_q$ are constants then so is $L_i(\gamma_1, \cdots, \gamma_q)$, $1 \leq i \leq r$, so that $L(\gamma_1, \cdots, \gamma_q) = 0$ if and only if $L_i(\gamma_1, \cdots, \gamma_q) = 0$, $1 \leq i \leq r$.

COROLLARY 2. *Let $\mathfrak{m}'$ be a set of polynomials in $\mathcal{F}[u_1, \cdots, u_q]$. There exists a set $\mathfrak{m}$ of polynomials in $\mathcal{C}[u_1, \cdots, u_q]$ such that q constants in an extension of $\mathcal{F}$ form a zero of $\mathfrak{m}'$ if and only if they form a zero of $\mathfrak{m}$.*

Proof. This follows from Corollary 1 since each polynomial in $u_1, \cdots, u_q$ is a linear combination of power products in $u_1, \cdots, u_q$.

COROLLARY 3. *Let $\gamma_1, \cdots, \gamma_q$ be constants in an extension of $\mathcal{F}$. Then $\partial^0\mathcal{F}\langle\gamma_1, \cdots, \gamma_q\rangle/\mathcal{F} = \partial^0\mathcal{C}\langle\gamma_1, \cdots, \gamma_q\rangle/\mathcal{C}$.*

Proof. By Corollary 2, $\gamma_{i_1}, \cdots, \gamma_{i_d}$ are algebraically dependent over $\mathcal{F}$ if and only if they are over $\mathcal{C}$.

COROLLARY 4. *If $\mathcal{G}$ is an extension of $\mathcal{F}$ with field of constants $\mathcal{D}$ and $\mathcal{E}$ is a differential subfield of $\mathcal{F}$ then $\mathcal{E}\langle\mathcal{D}\rangle \cap \mathcal{F} = \mathcal{E}\langle\mathcal{C}\rangle$.*

Proof. If $\alpha \in \mathcal{E}\langle\mathcal{C}\rangle$ then obviously $\alpha \in \mathcal{E}\langle\mathcal{D}\rangle \cap \mathcal{F}$. Conversely, let $\alpha \in \mathcal{E}\langle\mathcal{D}\rangle \cap \mathcal{F}$. Then there exist elements $e_1, \cdots, e_r \in \mathcal{E}$ linearly independent over constants and elements $d_1, \cdots, d_r, d'_1, \cdots, d'_r \in \mathcal{D}$ with $d'_1, \cdots, d'_r$ not all 0 (so that $\sum d'_ie_i \neq 0$) such that $\alpha = \sum d_ie_i/\sum d'_ie_i$, that is $\sum d'_ie_i\alpha - \sum d_ie_i = 0$. This means that $e_1\alpha, \cdots, e_r\alpha, e_1, \cdots, e_r$ are linearly dependent over constants, and therefore over $\mathcal{C}$; thus there exist elements $c_1, \cdots, c_r, c'_1, \cdots, c'_r \in \mathcal{C}$ not all 0 such that $\sum c'_ie_i\alpha - \sum c_ie_i = 0$. Since $e_1, \cdots, e_r$ are linearly independent over constants it follows that $c'_1, \cdots, c'_r$ are not all 0, so that $\sum c'_ie_i \neq 0$, whence $\alpha = \sum c_ie_i/\sum c'_ie_i \in \mathcal{E}\langle\mathcal{C}\rangle$.

Corollary 5. *Let $\mathfrak{D}$ be a field of constants containing $\mathcal{C}$ and contained in some extension of $\mathcal{F}$. Then the field of constants of $\mathcal{F}\langle\mathfrak{D}\rangle$ is $\mathfrak{D}$.*

Proof. Let e be a nonzero constant in $\mathcal{F}\langle\mathfrak{D}\rangle$. Then we may write $e\sum_{j=1}^{s}\alpha''_j d''_j=\sum_{i=1}^{r}\alpha'_i d'_i$, where $d'_i, d''_j \in \mathfrak{D}$, $\alpha'_i, \alpha''_j \in \mathcal{F}$, $\sum_{j=1}^{s}\alpha''_j d''_j \neq 0$. We suppose that of all such equations ours is one for which s is minimal, and also, without loss of generality, that $\alpha''_s=1$. For each $k (1 \leq k \leq m)$, $e\sum_{j=1}^{s-1}(\delta_k\alpha''_j)d''_j=\sum_{i=1}^{r}(\delta_k\alpha'_i)d'_i$, and this would contradict the minimal nature of s unless $\sum_{j=1}^{s}(\delta_k\alpha''_j)d''_j=0$; thus $\sum_{j=1}^{s}\alpha''_j d''_j$ is a constant. It follows that our corollary will be proved if we show that every nonzero constant $a \in \mathcal{F}\langle\mathfrak{D}\rangle$ of the form $a=\sum_{i=1}^{q}\alpha_i d_i$ $(\alpha_i \in \mathcal{F}, d_i \in \mathfrak{D})$ belongs to $\mathfrak{D}$. To this end we may suppose that $\alpha_1, \cdots, \alpha_q$ are linearly independent over constants and that every $d_i \neq 0$. Since $\sum_{i=1}^{q}(\delta_k\alpha_i)d_i=\delta_k a=0$ $(1 \leq k \leq m)$ it follows from Corollary 1 that there exist elements $c_1, \cdots, c_q$ of $\mathcal{C}$ not all 0 such that $\sum_{i=1}^{q}(\delta_k\alpha_i)c_i=0$ $(1 \leq k \leq m)$, so that the element $c=\sum_{i=1}^{q}\alpha_i c_i$ belongs to $\mathcal{C}$. Now, $\alpha_1, \cdots, \alpha_q$ are linearly independent and

$$\sum_{i=1}^{q}\alpha_i(ac_i-cd_i)=a\sum_{i=1}^{q}\alpha_i c_i-c\sum_{i=1}^{q}\alpha_i d_i=0,$$

so that each $ac_i-cd_i=0$, whence $a \in \mathfrak{D}$.

5. Universal extensions. Let $\mathcal{F}^*$ be a differential field and let $\mathcal{F}$ be a differential subfield of $\mathcal{F}^*$. We shall call $\mathcal{F}^*$ a *universal extension* of $\mathcal{F}$ if, for every finitely generated differential field extension $\mathcal{F}_1$ of $\mathcal{F}$ with $\mathcal{F}_1 \subseteq \mathcal{F}^*$ and every integer $n>0$ and every prime differential ideal Π of $\mathcal{F}_1\{y_1, \cdots, y_n\}$ not containing 1, there exists a generic zero $(\eta_1, \cdots, \eta_n)$ of Π with $\eta_1, \cdots, \eta_n \in \mathcal{F}^*$. A necessary and sufficient condition for an extension $\mathcal{F}^*$ of $\mathcal{F}$ to be universal is that for every finitely generated extension $\mathcal{F}_1$ of $\mathcal{F}$ with $\mathcal{F}_1 \subseteq \mathcal{F}^*$ and every finitely generated extension $\mathcal{G}$ of $\mathcal{F}_1$ there exist an isomorphim of $\mathcal{G}$ over $\mathcal{F}_1$ into $\mathcal{F}^*$ (that is, an isomorphism σ of $\mathcal{G}$ into $\mathcal{F}^*$ such that $\sigma a=a$ for every $a \in \mathcal{F}_1$). If $\mathcal{F}^*$ is a universal extension of $\mathcal{F}$ then $\mathcal{F}^*$ is a universal extension of every finitely generated extension of $\mathcal{F}$ contained in $\mathcal{F}^*$, and $\mathcal{F}^*$ is also a universal extension of every differential subfield of $\mathcal{F}$. If $\mathcal{F}^*$ is a universal extension of $\mathcal{F}$ then the degree of differential transcendence of $\mathcal{F}^*$ over $\mathcal{F}$ is infinite,

and (because of the Ritt basis theorem) for every integer $n > 0$ every differential ideal in $\mathcal{F}^*\{y_1, \cdots, y_n\}$ not containing 1 has a zero $(\eta_1, \cdots, \eta_n)$ with $\eta_1, \cdots, \eta_n \in \mathcal{F}^*$; in particular $\mathcal{F}^*$ is algebraically closed, and the field of constants of $\mathcal{F}^*$ is an alegbraically closed extension of the field of constants of $\mathcal{F}$ of infinite degree of transcendence.

We shall prove below that every differential field $\mathcal{F}$ has a universal extension $\mathcal{F}^*$. Once this fact is known it is possible to define the *manifold* of a set Φ of differential polynomials in $\mathcal{F}\{y_1, \cdots, y_n\}$ as the set of all zeros $(\eta_1, \cdots, \eta_n)$ of Φ with $\eta_1, \cdots, \eta_n \in \mathcal{F}^*$; this use of universal extensions extends the well-known procedure of modern algebraic geometry, and gives a workable definition of manifold free of the logical difficulty involved in using "the set of all extension of $\mathcal{F}$" (see Ritt [8], footnote 2 on p. 21).

Let Π be a prime differential ideal of $\mathcal{F}\{y_1, \cdots, y_n\}$. If $\mathcal{G}$ is an extension of $\mathcal{F}$ the ideal $\mathcal{G} \cdot \Pi$ of $\mathcal{G}\{y_1, \cdots, y_n\}$ is a perfect differential ideal (Proposition 1); we shall say that Π is *absolutely prime* if $\mathcal{G} \cdot \Pi$ is prime for every extension $\mathcal{G}$. If $\mathcal{G} \cdot \Pi$ is prime when we take for $\mathcal{G}$ some algebraically closed extension of $\mathcal{F}$ then Π is absolutely prime (because of Proposition 1 and the fact that every polynomial R over $\mathcal{F}$ which is irreducible over an algebraically closed extension of $\mathcal{F}$ is absolutely irreducible). In particular, if $\mathcal{F}$ is algebraically closed then every prime differential ideal of $\mathcal{F}\{y_1, \cdots, y_n\}$ is absolutely prime.

PROPOSITION 4. *Let I be a nonempty set of indices; for each $i \in I$ let n_i be an integer > 0; let $(y_{ij})_{i \in I, 1 \leq j \leq n_i}$ be a family of indeterminates; for each $i \in I$ let Π_i be an absolutely prime differential ideal of $\mathcal{F}\{y_{i1}, \cdots, y_{in_i}\}$ not containing* 1. *Then the ideal Π generated by $\bigcup_{i \in I} \Pi_i$ in $\mathcal{F}\{(y_{ij})_{i \in I, 1 \leq j \leq n_i}\}$ is a prime differential ideal not containing* 1. *If I is finite then Π is absolutely prime and* $\operatorname{ord} \Pi = \sum_{i \in I} \operatorname{ord} \Pi_i$.

Proof. That Π is a differential ideal is obvious. To prove that Π is prime and $1 \notin \Pi$ it suffices to consider the case in which I is finite; by induction then the entire proposition can be reduced to the case in which I consists of two elements. Accordingly, let I consist of the numbers 1 and 2. Then Π consists of all differential polynomials P which can be writen in the form

$$(2) \qquad P = \sum_{k_1} C_{2k_1} P_{1k_1} + \sum_{k_2} C_{1k_2} P_{2k_2} (P_{ik_i} \in \Pi_i, C_{ik} \in \mathcal{F}\{y_{i1}, \cdots, y_{in}\}).$$

Let $(\eta_{11}, \cdots, \eta_{1n_1})$ be a generic zero of Π_1. Since Π_2 is absolutely prime the ideal Λ_2 generated by Π_2 in $\mathcal{F}\langle\eta_{11}, \cdots, \eta_{1n_1}\rangle\{y_{21}, \cdots, y_{2n_2}\}$ is a prime differential ideal, and obviously $1 \notin \Lambda_2$; let $(\eta_{21}, \cdots, \eta_{2n_2})$ be a generic zero

of Λ_2. We shall show that $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$ is a generic zero of Π, thereby proving that Π is prime, that $1 \notin \Pi$, and that

$$\begin{aligned}\operatorname{ord} \Pi &= \partial^0 \mathcal{F}\langle \eta_{11}, \cdots, \eta_{1n}, \eta_{21}, \cdots, \eta_{2n_2}\rangle/\mathcal{F} \\ &= \partial^0 \mathcal{F}\langle \eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2}\rangle/\mathcal{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle \\ &\quad + \partial^0 \mathcal{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle/\mathcal{F} = \operatorname{ord} \Lambda_2 + \operatorname{ord} \Pi_1 = \operatorname{ord} \Pi_2 + \operatorname{ord} \Pi_1.\end{aligned}$$

It is clear from (2) that $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$ is a zero of Π. Let P be any differential polynomial in $\mathcal{F}\{y_{11}, \cdots, y_{1n_1}, y_{21}, \cdots, y_{2n_2}\}$ which vanishes at $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$. Then

$$P(\eta_{11}, \cdots, \eta_{1n_1}, y_{21}, \cdots, y_{2n_2}) \in \Lambda_2.$$

We now write

$$P(\eta_{11}, \cdots, \eta_{1n_1}, y_{21}, \cdots, y_{2n_2}) = \sum_{k_2} C_{1k_2}(\eta_{11}, \cdots, \eta_{1n_1}) P_{2k_2}(y_{21}, \cdots, y_{2n_2}),$$

where

$$C_{1k_2}(y_{11}, \cdots, y_{1n_1}) \in \mathcal{F}\{y_{11}, \cdots, y_{1n_1}\},\ P_{2k_2}(y_{21}, \cdots, y_{2n_2}) \in \mathcal{F}\{y_{21}, \cdots, y_{2n_2}\},$$

and the elements $C_{1k_2}(\eta_{11}, \cdots, \eta_{1n_1})$ of $\mathcal{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle$ are linearly independent over $\mathcal{F}$; it is easy to see, since $\Lambda_2 = \mathcal{F}\langle \eta_{11}, \cdots, \eta_{1n_1}\rangle \cdot \Pi_2$, that each $P_{2k_2} \in \Pi_2$. Let

$$Q = P - \sum_{k_2} C_{1k_2} P_{2k_2},$$

so that $Q(\eta_{11}, \cdots, \eta_{1n_1}, y_{21}, \cdots, y_{2n_2}) = 0$; if we write

$$Q = \sum_{k_1} C_{2k_1} P_{1k_1},$$

where the C_{2k_1} are distinct power products in $y_{21}, \cdots, y_{2n_2}$ and their derivatives of various orders, and each $P_{1k_1} \in \mathcal{F}\{y_{11}, \cdots, y_{1n_1}\}$, then each $P_{1k_1}(\eta_{11}, \cdots, \eta_{1n_1}) = 0$, so that each $P_{1k_1} \in \Pi_1$. It follows that P can be written in the form (2) and therefore belongs to Π; therefore $(\eta_{11}, \cdots, \eta_{1n_1}, \eta_{21}, \cdots, \eta_{2n_2})$ is a generic zero of Π.

To complete the proof it remains to show that Π is absolutely prime. To this end let $\mathcal{G}$ be any extension of $\mathcal{F}$. Clearly $\mathcal{G} \cdot \Pi$ is the ideal generated by $(\mathcal{G} \cdot \Pi_1) \cup (\mathcal{G} \cdot \Pi_2)$ in $\mathcal{G}\{y_{11}, \cdots, y_{1n_1}, y_{21}, \cdots, y_{2n_2}\}$, and $\mathcal{G} \cdot \Pi$, $\mathcal{G} \cdot \Pi_2$ are prime (because Π_1, Π_2 are absolutely prime); therefore by what we have already proved $\mathcal{G} \cdot \Pi$ is prime. Thus Π is absolutely prime.

REMARK. We observe from the proof that the hypothesis in Proposition 4 that each Π_i be absolutely prime may be weakened. It is enough to assume that, for each $i \in I$, Π_i is prime and $\mathcal{F}_i \cdot \Pi_i$ is prime whenever $\mathcal{F}_i$ is an

extension of $\mathcal{F}$ obtained by the adjunction of generic zeros of a finite number of Π_j's with $j \neq i$. Except for the statement that Π is *absolutely* prime, the conclusion of Proposition 4 is then valid (Π still being prime).

THEOREM. *Every differential field has a universal extension.*

Proof. We lose no generality in assuming that the given differential field $\mathcal{F}$ is algebraically closed, for a universal extension of an algebraic closure of $\mathcal{F}$ is a universal extension of $\mathcal{F}$. We shall show that for every algebraically closed differential field $\mathcal{G}$ there exists an extension $\mathcal{G}^\dagger$ of $\mathcal{G}$ with the following two properties: 1) $\mathcal{G}^\dagger$ is algebraically closed; 2) for every integer $n > 0$ and for every prime differential ideal Π of $\mathcal{G}\{y_1, \cdots, y_n\}$ not containing 1 there exists a generic zero $(\eta_1, \cdots, \eta_n)$ of Π with $\eta_1, \cdots, \eta_n \in \mathcal{G}^\dagger$. Once this is done we can define inductively a sequence of differential fields $\mathcal{F}^{(k)}$ such that $\mathcal{F}^{(0)} = \mathcal{F}$ and $\mathcal{F}^{(k+1)} = \mathcal{F}^{(k)\dagger}$ for every integer $k \geq 0$; the union $\mathcal{F}^* = \bigcup \mathcal{F}^{(k)}$ will then be a differential field which, as is easy to see, is a universal extension of $\mathcal{F}$.

Let $\mathfrak{P}_n$ be the set of all prime differential ideals in $\mathcal{G}\{y_1, \cdots, y_n\}$ which do not contain 1; since $\mathcal{G}$ is algebraically closed, every element of $\mathfrak{P}_n$ is absolutely prime. Let $(y_{n\Pi j})_{1 \leq n < \infty, \Pi \in \mathfrak{P}_n, 1 \leq j \leq n}$ be a family of indeterminates. For each $\Pi \in \mathfrak{P}_n$ let $\Lambda(n, \Pi)$ denote the set which is obtained when in all the differential polynomials in Π we replace y_j by $y_{n\Pi j}$ $(1 \leq j \leq n)$; $\Lambda(n, \Pi)$ is obviously an absolutely prime differential ideal of $\mathcal{G}\{y_{n\Pi 1}, \cdots, y_{n\Pi n}\}$ which does not contain 1. It follows from Proposition 4 that the ideal Λ generated by $\bigcup_{1 \leq n < \infty, \Pi \in \mathfrak{P}_n} \Lambda(n, \Pi)$ in the differential ring $\mathcal{R} = \mathcal{G}\{(y_{n\Pi j})_{1 \leq n < \infty, \Pi \in \mathfrak{P}_n, 1 \leq j \leq n}\}$ is a prime differential ideal not containing 1. The differential ring of residue classes $\mathcal{R}/\Lambda$ is therefore a differential domain of integrity, which can be embedded in its differential field of quotients $\mathcal{G}'$. Since $1 \notin \Lambda$, the canonical homomorphism h of $\mathcal{R}$ onto $\mathcal{R}/\Lambda$ maps $\mathcal{G}$ isomorphically; therefore we may identify each element $a \in \mathcal{G}$ with its image $h(a) \in \mathcal{G}'$. With this identification $\mathcal{G}$ becomes a differential subfield of $\mathcal{G}'$. It is now easy to see that if we set $\eta_{n\Pi j} = h(y_{n\Pi j})$ for all n, Π, j then, for each n and each $\Pi \in \mathfrak{P}_n$, $(\eta_{n\Pi 1}, \cdots, \eta_{n\Pi n})$ is a generic zero of $\Lambda(n, \Pi)$, and consequently a generic zero of Π. Therefore if we let $\mathcal{G}^\dagger$ be an algebraic closure of $\mathcal{G}'$ then $\mathcal{G}^\dagger$ will have the required properties 1), 2) above. As we have seen, this suffices to prove the theorem.

Chapter II. Algebraic groups of automorphisms.

Throughout the rest of this paper $\mathcal{G}$ will denote a differential field with algebraically closed field of constants $\mathcal{C}$, and $\mathcal{F}$ will denote a differential subfield of $\mathcal{G}$, with the same field of constants $\mathcal{C}$, such that $\mathcal{G}$ is finitely generated and of finite transcendence degree over $\mathcal{F}$. The relative algebraic closure of $\mathcal{F}$ in $\mathcal{G}$ will be denoted by $\mathcal{F}^0$. All differential fields mentioned will tacitly be assumed to lie in a universal extension of $\mathcal{G}$ fixed once and for all; in particular, every isomorphism of $\mathcal{G}$ will be an isomorphism into this universal extension. The identity isomorphism of $\mathcal{G}$ will be denoted by ι. The field of constants of the universal extension will be denoted by $\mathcal{C}^*$.

1. Specializations of isomorphisms. Let $\sigma_1, \cdots, \sigma_p, \tau_1, \cdots, \tau_p$ be isomorphisms of $\mathcal{G}$; we shall say that $(\tau_1, \cdots, \tau_p)$ *is a specialization* of $\sigma_1, \cdots, \sigma_p)$ if $(\tau_i\alpha)_{1 \leq i \leq p, \alpha \in G}$ is a specialization of $(\sigma_i\alpha)_{1 \leq i \leq p, \alpha \in G}$ over $\mathcal{G}$. If $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ such that $(\sigma_1, \cdots, \sigma_p)$ is a specialization of $(\tau_1, \cdots, \tau_p)$ the we shall say that $(\tau_1, \cdots, \tau_p)$ is a *generic* specialization of $(\sigma_1, \cdots, \sigma_p)$. A specialization which is not generic will be called *nongeneric*. The relation "τ is a generic specialization of σ" is an equivalence on the set of all isomorphisms of $\mathcal{G}$, and two isomorphisms of $\mathcal{G}$ which are in this relation will accordingly be called *equivalent*.

2. Isolated isomorphisms. We shall say that σ is an *isolated* isomorphism of $\mathcal{G}$ over $\mathcal{F}$ if σ is an isomorphism of $\mathcal{G}$ over $\mathcal{F}$ such that there does not exist an isomorphism of $\mathcal{G}$ over $\mathcal{F}$ of which σ is a nongeneric specialization.

Let $\eta_1, \cdots, \eta_n$ be elements such that $\mathcal{G} = \mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle$. If $\sigma_1, \cdots, \sigma_p$, $\tau_1, \cdots, \tau_p$ are isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ then $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ if and only if $(\tau_i\eta_j)_{1 \leq i \leq p, 1 \leq j \leq n}$ is a specialization of $(\tau_i\eta_j)_{1 \leq i \leq p, 1 \leq j \leq n}$ over $\mathcal{G}$.

Let Π be the prime differential ideal of $\mathcal{F}\{y_1, \cdots, y_n\}$ with generic zero $(\eta_1, \cdots, \eta_n)$, so that (Chapter I, Proposition 1) $\mathcal{G} \cdot \Pi$ is a perfect differential ideal of $\mathcal{G}\{y_1, \cdots, y_n\}$, and let $\Pi_1, \cdots, \Pi_h$ be the minimal prime differential ideal divisors of $\mathcal{G} \cdot \Pi$. Let $(\eta_{i1}, \cdots, \eta_{in})$ be a generic zero of Π_i; then $(\eta_{i1}, \cdots, \eta_{in})$ is also a generic zero of Π, so that there exists a unique isomorphism χ_i of $\mathcal{G}$ over $\mathcal{F}$ such that $\chi_i\eta_j = \eta_{ij}$ $(1 \leq j \leq n)$. It is obvious that if $i \neq i'$ then χ_i is not equivalent to $\chi_{i'}$. If σ is any isomorphism of $\mathcal{G}$ over $\mathcal{F}$ then $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Π, so that (Chapter I,

Proposition 1) $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a zero of precisely one Π_i; therefore σ is a specialization of precisely one χ_i. We have thus proved the following result.

PROPOSITION 1. *$\chi_1, \cdots, \chi_n$ are inequivalent isolated isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, and every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is a specialization of precisely one of these.*

By Proposition 1 an isomorphism σ of $\mathcal{G}$ over $\mathcal{F}$ is isolated if and only if σ is equivalent to χ_i for some i, that is (Chapter I, Proposition 2) if and only if

$$\begin{aligned}\partial^0\mathcal{G}\langle\sigma\mathcal{G}\rangle/\mathcal{G} &= \partial^0\mathcal{G}\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle/\mathcal{G}\\ &= \partial^0\mathcal{G}\langle\chi_i\eta_1, \cdots, \chi_i\eta_n\rangle/\mathcal{G} = \operatorname{ord}\Pi_i = \operatorname{ord}\Pi = \partial^0\mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle/\mathcal{F}\\ &= \partial^0\mathcal{G}/\mathcal{F}.\end{aligned}$$

Since $\mathcal{G}\langle\sigma\mathcal{G}\rangle = (\sigma\mathcal{G})\langle\mathcal{G}\rangle$ and $\partial^0\mathcal{G}/\mathcal{F} = \partial^0(\sigma\mathcal{G})/\mathcal{F}$ it follows that σ is an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ if and only if σ^{-1} is an isolated isomorphism of $\sigma\mathcal{G}$ over $\mathcal{F}$. Thus we have the following result.

PROPOSITION 2. *Let σ be an isomorphism of $\mathcal{G}$ over $\mathcal{F}$. The following three statements are equivalent:* 1) *σ is an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$;* 2) *σ^{-1} is an isolated isomorphism of $\sigma\mathcal{G}$ over $\mathcal{F}$;* 3) $\partial^0\mathcal{G}\langle\sigma\mathcal{G}\rangle/\mathcal{G} = \partial^0\mathcal{G}/\mathcal{F}$.

Now let σ be any isomorphism of $\mathcal{G}$ over $\mathcal{F}$, and suppose that σ leaves invariant some element $\zeta \in \mathcal{G}$ which is transcendental over $\mathcal{F}$. Let Λ be the prime differential ideal of $\mathcal{F}\{z, y_1, \cdots, y_n\}$ with generic zero $(\zeta, \eta_1, \cdots, \eta_n)$ and let $\Lambda_1, \cdots, \Lambda_k$ be the minimal prime differential ideal divisors of the perfect differential ideal $\mathcal{G}\cdot\Lambda$ of $\mathcal{G}\{z, y_1, \cdots, y_n\}$. $(\zeta, \sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Λ and therefore a zero of Λ_i for some i; let $(\zeta', \eta'_1, \cdots, \eta'_n)$ be a generic zero of Λ_i. Then $(\zeta', \eta'_1, \cdots, \eta'_n)$ is a generic zero of Λ, so that there exists a unique isomorphism σ' of $\mathcal{G}$ over $\mathcal{F}$ such that $\sigma'\zeta = \zeta'$, $\sigma'\eta_1 = \eta'_1, \cdots, \sigma'\eta_n = \eta'_n$; it is clear that σ is a specialization of σ'. If ζ were invariant under σ', that is if we had $\zeta' = \zeta$, then we would have (because ζ is transcendental over $\mathcal{F}$)

$$\begin{aligned}&\partial^0\mathcal{G}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathcal{G}\\ &\qquad \leq \partial^0\mathcal{F}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathcal{F}\langle\zeta'\rangle < \partial^0\mathcal{F}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathcal{F}\langle\zeta'\rangle\\ &\qquad\qquad + \partial^0\mathcal{F}\langle\zeta'\rangle/\mathcal{F} = \partial^0\mathcal{F}\langle\zeta', \eta'_1, \cdots, \eta'_n\rangle/\mathcal{F},\end{aligned}$$

or in other words $\operatorname{ord}\Lambda_i < \operatorname{ord}\Lambda$, contradicting Chapter I, Proposition 1. Therefore $\sigma'\zeta \neq \zeta$; since $\sigma\zeta = \zeta$ this means that σ is a nongeneric specialization of σ', so that σ can not be an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$.

This shows that the field of invariants of an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ must be contained in $\mathcal{F}^0$.

If an element of $\mathcal{G}$ is invariant under every isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ (or, equivalently, under the isolated isomorphisms $\chi_1, \cdots, \chi_h$) then the element is invariant under every isomorphism of $\mathcal{G}$ over $\mathcal{F}$, and therefore (Kolchin [3], § 12) belongs to $\mathcal{F}$.

Consider again the prime differential ideal Π of $\mathcal{F}\{y_1, \cdots, y_n\}$ with generic zero $(\eta_1, \cdots, \eta_n)$. It is a consequence of Chapter I, Proposition 1 that the minimal prime differential ideal divisors of the perfect differential ideal $\mathcal{F}^0 \cdot \Pi$ of $\mathcal{F}^0\{y_1, \cdots, y_n\}$ are h in number, one being contained in and generating each of the minimal prime differential ideal divisors $\Pi_1, \cdots, \Pi_h$ of $\mathcal{G} \cdot \Pi$; we denote the minimal prime differential ideal divisor of $\mathcal{F}^0 \cdot \Pi$ which is contained in Π_i by $\Pi_i{}^0$, so that $\mathcal{G} \cdot \Pi_i{}^0 = \Pi_i$. Now $(\eta_1, \cdots, \eta_n)$ is a zero of precisely one $\Pi_i{}^0$, say of $\Pi_1{}^0$; because

$$\partial^0 \mathcal{F}^0\langle\eta_1, \cdots, \eta_n\rangle/\mathcal{F}^0 = \partial^0 \mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle/\mathcal{F} = \operatorname{ord} \Pi = \operatorname{ord} \Pi_1$$

we see (Chapter I, Proposition 2) that $(\eta_1, \cdots, \eta_n)$ is a generic zero of $\Pi_1{}^0$. Also, the identity isomorphism ι is a specialization of precisely one χ_i. Because $(\eta_1, \cdots, \eta_n) = (\iota\eta_1, \cdots, \iota\eta_n)$ is a zero of $\Pi_1{}^0$ and $\Pi_1 = \mathcal{G} \cdot \Pi_1{}^0$, ι must be a specialization of χ_1. It follows from this that the restriction of χ_1 to $\mathcal{F}^0$ is the identity isomorphism of $\mathcal{F}^0$. If σ is any isomorphism of $\mathcal{G}$ such that the restriction of σ to $\mathcal{F}^0$ is the identity then $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a zero of $\Pi_1{}^0$ (because $(\eta_1, \cdots, \eta_n)$ is) and consequently a zero of Π_1, so that σ is a specialization of χ_1.

Collecting these remarks we have the following result.

PROPOSITION 3. *The field of invariants of an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is contained in $\mathcal{F}^0$; the field of invariants of a complete set of representatives of the h equivalence classes of isolated isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ is $\mathcal{F}$. If σ_0 is an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}$ of which ι is a specialization then the field of invariants of σ_0 is $\mathcal{F}^0$; every isomorphism of $\mathcal{G}$ which leaves all the elements of $\mathcal{F}^0$ invariant is a specialization of σ_0.*

3. Strong isomorphisms. If σ is any isomorphism of $\mathcal{G}$ we denote the field of constants of $\mathcal{G}\langle\sigma\mathcal{G}\rangle$ by $\mathcal{C}_\sigma$.

We shall say that an isomorphism σ of $\mathcal{G}$ is *strong* if

$$\sigma\mathcal{G} \subset \mathcal{G}\langle\mathcal{C}^*\rangle, \ \mathcal{G} \subset (\sigma\mathcal{G})\langle\mathcal{C}^*\rangle.$$

By Chapter I, Corollary 4 to Proposition 3 (with $\mathcal{G}$, $\mathcal{G}\langle\sigma\mathcal{G}\rangle$, $\mathcal{G}\langle\mathcal{C}^*\rangle$ playing the role of $\mathcal{E}$, $\mathcal{F}$, $\mathcal{G}$ in that Corollary) the first of these inclusions is

equivalent to $\sigma\mathcal{G} \subset \mathcal{G}\langle\mathcal{C}_\sigma\rangle$; similarly, the second of these inclusions is equivalent to $\mathcal{G} \subset (\sigma\mathcal{G})\langle\mathcal{C}_\sigma\rangle$. Consequently σ is strong if and only if

$$\mathcal{G}\langle\mathcal{C}_\sigma\rangle = \mathcal{G}\langle\sigma\mathcal{G}\rangle = (\sigma\mathcal{G})\langle\mathcal{C}_\sigma\rangle. \tag{1}$$

Obviously every automorphism of $\mathcal{G}$ is strong.

Let $\Pi, \eta_1, \cdots, \eta_n$ have the same significance as in § 2. Because the field of constants $\mathcal{C}$ of $\mathcal{F}$ is algebraically closed it is easy to see that every polynomial irreducible over $\mathcal{F}$ remains irreducible over $\mathcal{F}\langle\mathcal{C}^*\rangle$. It follows (Chapter I, Proposition 1) that the differential ideal $\Pi^* = \mathcal{F}\langle\mathcal{C}^*\rangle \cdot \Pi$ of $\mathcal{F}\langle\mathcal{C}^*\rangle\{y_1, \cdots, y_n\}$ is prime, and $\operatorname{ord}\Pi^* = \operatorname{ord}\Pi$.

Let σ be any strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$; $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Π, and therefore a zero of Π^*. For every finite subset $\mathfrak{c}$ of $\mathcal{C}^*$ we have (Chapter I, Corollary 3 to Proposition 3)

$$\begin{aligned}\partial^0\mathcal{F}\langle\mathfrak{c}\rangle\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle/\mathcal{F}\langle\mathfrak{c}\rangle &= \partial^0(\sigma\mathcal{G})\langle\mathfrak{c}\rangle/\mathcal{F}\langle\mathfrak{c}\rangle \\ &= \partial^0(\sigma\mathcal{G})\langle\mathfrak{c}\rangle/\mathcal{F} - \partial^0\mathcal{F}\langle\mathfrak{c}\rangle/\mathcal{F} = \partial^0(\sigma\mathcal{G}\langle\mathfrak{c}\rangle/\sigma\mathcal{G} + \partial^0(\sigma\mathcal{G})/\mathcal{F} \\ &\quad - \partial^0\mathcal{F}\langle\mathfrak{c}\rangle/\mathcal{F} = \partial^0\mathcal{C}\langle\mathfrak{c}\rangle/\mathcal{C} + \operatorname{ord}\Pi - \partial^0\mathcal{C}\langle\mathfrak{c}\rangle/\mathcal{C} = \operatorname{ord}\Pi^*;\end{aligned}$$

since this holds for every finite subset $\mathfrak{c}$ of $\mathcal{C}^*$ we infer that

$$\partial^0\mathcal{F}\langle\mathcal{C}^*\rangle\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle/\mathcal{F}\langle\mathcal{C}^*\rangle = \operatorname{ord}\Pi^*,$$

so that (Chapter I, Corollary to Proposition 2) $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic zero of Π^*. In the same way we also show that $(\eta_1, \cdots, \eta_n)$ is a general zero of Π^*, so that $(\sigma\eta_1, \cdots, \sigma\eta_n)$ is a generic specialization of $(\eta_1, \cdots, \eta_n)$ over $\mathcal{F}\langle\mathcal{C}^*\rangle$. Therefore there exists a unique isomorphism σ^* of $\mathcal{F}\langle\mathcal{C}^*\rangle\langle\eta_1, \cdots, \eta_n\rangle$ over $\mathcal{F}\langle\mathcal{C}^*\rangle$ onto $\mathcal{F}\langle\mathcal{C}^*\rangle\langle\sigma\eta_1, \cdots, \sigma\eta_n\rangle$ such that $\sigma^*\eta_1 = \sigma\eta_1, \cdots, \sigma^*\eta_n = \sigma\eta_n$, that is there exists a unique isomorphism σ^* of $\mathcal{G}\langle\mathcal{C}^*\rangle$ over $\mathcal{F}\langle\mathcal{C}^*\rangle$ which extends σ. Since by (1)

$$\sigma^*(\mathcal{G}\langle\mathcal{C}^*\rangle) = (\sigma\mathcal{G})\langle\mathcal{C}^*\rangle = (\sigma\mathcal{G})\langle\mathcal{C}_\sigma\rangle\langle\mathcal{C}^*\rangle = \mathcal{G}\langle\mathcal{C}_\sigma\rangle\langle\mathcal{C}^*\rangle = \mathcal{G}\langle\mathcal{C}^*\rangle,$$

we see that σ^* is an automorphism of $\mathcal{G}\langle\mathcal{C}^*\rangle$.

Now let us start at the other end with any automorphism σ^* of $\mathcal{G}\langle\mathcal{C}^*\rangle$ over $\mathcal{F}\langle\mathcal{C}^*\rangle$. The restriction σ of σ^* to $\mathcal{G}$ is then an isomorphism of $\mathcal{G}$ over $\mathcal{F}$. Obviously $\sigma\mathcal{G} \subset \mathcal{G}\langle\mathcal{C}^*\rangle$ and $\mathcal{G} \subset (\sigma\mathcal{G})\langle\mathcal{C}^*\rangle$, so that σ is strong. We have thus proved the following result.

PROPOSITION 4. *The mapping which to each automorphism of $\mathcal{G}\langle\mathcal{C}^*\rangle$ over $\mathcal{F}\langle\mathcal{C}^*\rangle$ assigns its restriction to $\mathcal{G}$ is one-to-one onto the set of all strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$.*

In virtue of Proposition 4 we may identify each strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$ with the automorphism of $\mathcal{G}\langle\mathcal{C}^*\rangle$ over $\mathcal{F}\langle\mathcal{C}^*\rangle$ of which it is the restriction. This identification permits us to multiply any two strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, and to consider the set of all of them as a group. We shall denote this group of all strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ by $\mathfrak{G}^*$, and the subgroup of $\mathfrak{G}^*$ consisting of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ by $\mathfrak{G}$.

If σ is a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$, application of σ^{-1} to (1) shows that $\mathcal{C}_\sigma \subseteq \mathcal{C}_{\sigma^{-1}}$; interchanging σ and σ^{-1} reverses the inclusion; we conclude that

$$\mathcal{C}_\sigma = \mathcal{C}_{\sigma^{-1}}. \tag{2}$$

4. Specializations of strong isomorphisms.

PROPOSITION 5. *A specialization of a strong isomorphism of $\mathcal{G}$ is always strong.*

Proof. Let σ be a strong isomorphism of $\mathcal{G}$. By (1), for each $\alpha \in \mathcal{G}$ we may write a relation $\sigma\alpha = \sum_{i=1}^{r} a_i\beta_i / \sum_{i=1}^{r} b_i\beta_i$, where a_i, b_i are constants, $\beta_1, \cdots, \beta_r$ are elements of $\mathcal{G}$ linearly independent over constants, and $\sum b_i\beta_i \neq 0$. Therefore $\beta_1, \cdots, \beta_r, \beta_1\sigma\alpha, \cdots, \beta_r\sigma\alpha$ are linearly dependent over constants so that (Chapter I, Proposition 3) the differential polynomial $W_{\theta_1\ldots\theta_{2r}}(\beta_1, \cdots, \beta_r, \beta_1 y, \cdots, \beta_r y) \in \mathcal{G}\{y\}$ vanishes at $\sigma\alpha$ for all choices of the differential operators $\theta_1, \cdots, \theta_{2r}$ of order $< 2r$. If σ' is a specialization of σ then this differential polynomial also vanishes at $\sigma'\alpha$ so that (Chapter I, Proposition 3) there exist elements $a'_1, \cdots, a'_r, b'_1, \cdots, b'_r \in \mathcal{C}_{\sigma'}$ not all 0 such that $\sum a'_i\beta_i - \sum b'_i\beta_i\sigma'\alpha = 0$. Since $\beta_1, \cdots, \beta_r$ are linearly independent over constants not every b'_i is 0 so that $\sum b'_i\beta_i \neq 0$ and

$$\sigma'\alpha = \sum a'_i\beta_i / \sum b'_i\beta_i \in \mathcal{G}\langle\mathcal{C}_{\sigma'}\rangle.$$

Thus $\mathcal{G}\langle\sigma'\mathcal{G}\rangle = \mathcal{G}\langle\mathcal{C}_{\sigma'}\rangle$.

Again by (1), for each $\alpha \in \mathcal{G}$ we may write $\alpha = \sum_{i=1}^{r} a_i\sigma\beta_i / \sum_{i=1}^{r} b_i\sigma\beta_i$, where a_i, b_i are constants, $\sigma\beta_1, \cdots, \sigma\beta_r$ are elements of $\sigma\mathcal{G}$ linearly independent over constants, and $\sum b_i\sigma\beta_i \neq 0$ (so that $\beta_1, \cdots, \beta_r$ are elements of $\mathcal{G}$ linearly independent over constants and $\sum b_i\beta_i \neq 0$). Therefore the differential polynomial $W_{\theta_1\ldots\theta_{2r}}(y_1, \cdots, y_r, \alpha y_1, \cdots, \alpha y_r) \in \mathcal{G}\{y_1, \cdots, y_r\}$ vanishes at $(\sigma\beta_1, \cdots, \sigma\beta_r)$ for every choice of $\theta_1, \cdots, \theta_{2r}$ of order $< 2r$, so that this differential polynomial also vanishes at $(\sigma'\beta_1, \cdots, \sigma'\beta_r)$. This implies that there exist constants $a'_i, b'_i \in \mathcal{C}_{\sigma'}$ not all 0 such that $\sum a'_i\sigma'\beta_i - \sum \alpha b'_i\sigma'\beta_i = 0$;

because $\beta_1, \cdots, \beta_r$ are linearly independent over constants $\sigma'\beta_1, \cdots, \sigma'\beta_r$ are too, so that $\sum b'_i\sigma'\beta_i \neq 0$. Therefore

$$\alpha = \sum a'_i\sigma'\beta_i / \sum b'_i\sigma'\beta_i \,\varepsilon\, (\sigma'\mathcal{G})\langle \mathcal{C}_{\sigma'}\rangle,$$

so that $\mathcal{G}\langle\sigma'\mathcal{G}\rangle = (\sigma'\mathcal{G})\langle\mathcal{C}_{\sigma'}\rangle$. It follows that σ' satisfies (1) and σ' is strong.

REMARK. We observe from the proof of Proposition 5 that if σ is an isomorphism of $\mathcal{G}$ which satisfies the first (second) equation (1) then every specialization of σ also satisfies the first (second) equation (1).

PROPOSITION 6. *If $\sigma_1, \cdots, \sigma_p$ are strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ and if $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ then $(\tau_1^{-1}, \tau_1^{-1}\tau_2, \cdots, \tau_1^{-1}\tau_p)$ is a specialization of $(\sigma_1^{-1}, \sigma_1^{-1}\sigma_2, \cdots, \sigma_1^{-1}\sigma_p)$.*

Proof. Let $F \,\varepsilon\, \mathcal{G}\{(z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\, 1\leq j\leq q}\}$; if we denote the coefficients in F by $\beta_1, \cdots, \beta_r$ then we may write

$$F((z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\, 1\leq j\leq q}) = G((z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\, 1\leq j\leq q}, (\beta_k)_{1\leq k\leq r}),$$

where $G \,\varepsilon\, \mathcal{F}\{(z_{1j})_{1\leq j\leq q}, (z_{ij})_{2\leq i\leq p,\, 1\leq j\leq q}, (y_k)_{1\leq k\leq r}\}$. If F vanishes at $((\sigma_1^{-1}\alpha_j)_{1\leq j\leq q}, (\sigma_1^{-1}\sigma_i\alpha_j)_{2\leq i\leq p,\, 1\leq j\leq q})$, that is if G vanishes at

$$((\sigma_1^{-1}\alpha_j)_{1\leq j\leq q}, (\sigma_1^{-1}\sigma_i\alpha_j)_{2\leq i\leq p,\, 1\leq j\leq q}, (\beta_k)_{1\leq k\leq r}),$$

then application of σ_1 shows that G vanishes at

$$((\alpha_j)_{1\leq j\leq q}, (\sigma_i\alpha_j)_{2\leq i\leq p,\, 1\leq j\leq q}, (\sigma_1\beta_k)_{1\leq k\leq r});$$

since $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$ this implies that G vanishes at

$$((\alpha_j)_{1\leq j\leq q}, (\tau_i\alpha_j)_{2\leq i\leq p,\, 1\leq j\leq q}, (\tau_1\beta_k)_{1\leq k\leq r});$$

application of τ_1^{-1} shows that G vanishes at

$$((\tau_1^{-1}\alpha_j)_{1\leq j\leq q}, (\tau_1^{-1}\tau_i\alpha_j)_{2\leq i\leq p,\, 1\leq j\leq q}, (\beta_k)_{1\leq k\leq r}),$$

that is that F vanishes at $((\tau_1^{-1}\alpha_j)_{1\leq j\leq q}, (\tau_1^{-1}\tau_i\alpha_j)_{2\leq i\leq p,\, 1\leq j\leq q})$. It follows that $\tau_1^{-1}, \tau_1^{-1}\tau_2, \cdots, \tau_1^{-1}\tau_p)$ is a specialization of $(\sigma_1^{-1}, \sigma_1^{-1}\sigma_2, \cdots, \sigma_1^{-1}\sigma_p)$.

Let $\sigma_1, \cdots, \sigma_p$ be isomorphisms of $\mathcal{G}$ which have the following property: whenever $\tau_1, \cdots, \tau_p$ are isomorphisms of $\mathcal{G}$ such that τ_i is a specialization of σ_i $(1 \leq i \leq p)$ then $(\tau_1, \cdots, \tau_p)$ is a specialization of $(\sigma_1, \cdots, \sigma_p)$; we shall say under these circumstances that $\sigma_1, \cdots, \sigma_p$ are *independent.*

PROPOSITION 7. *The strong isomorphisms $\sigma_1, \cdots, \sigma_p$ of $\mathcal{G}$ over $\mathcal{F}$ are independent if and only if $\partial^0 \mathcal{G}\langle\sigma_1\mathcal{G}, \cdots, \sigma_p\mathcal{G}\rangle/\mathcal{G} = \sum_{i=1}^{p} \partial^0 \mathcal{G}\langle\sigma_i\mathcal{G}\rangle/\mathcal{G}$.*

Proof. Let Λ_i denote the prime differential ideal in $\mathcal{G}\{y_{i1}, \cdots, y_{im}\}$ with generic zero $(\sigma_i\eta_i, \cdots, \sigma_i\eta_n)$, where as before $\eta_1, \cdots, \eta_n$ generate $\mathcal{G} : \mathcal{G} = \mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle$. If $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic zero of Λ_i then $(\eta'_{i1}, \cdots, \eta'_{in})$ is a generic specialization of $(\sigma_i\eta_1, \cdots, \sigma_i\eta_n)$ over $\mathcal{G}$, so that there exists an isomorphism of $\mathcal{G}\langle\sigma_i\mathcal{G}\rangle = \mathcal{G}\langle\sigma_i\eta_1, \cdots, \sigma_i\eta_n\rangle$ onto $\mathcal{G}\langle\eta'_{i1}, \cdots, \eta'_{in}\rangle$ over $\mathcal{G}$; since $\mathcal{G}\langle\sigma_i\mathcal{G}\rangle = \mathcal{G}\langle\mathcal{C}_{\sigma_i}\rangle$ we have $\mathcal{G}\langle\eta'_{i1}, \cdots, \eta'_{in}\rangle = \mathcal{G}\langle\mathcal{C}'_i\rangle$, where $\mathcal{C}'_i$ is the field of constants of $\mathcal{G}\langle\eta'_{i1}, \cdots, \eta'_{in}\rangle$. The field of constants of $\mathcal{G}$ is $\mathcal{C}$, which is algebraically closed; it easily follows that every polynomial which is irreducible over $\mathcal{G}$ remains irreducible over

$$\mathcal{G}_{i_0} = \mathcal{G}\langle\mathcal{C}'_i, \cdots, \mathcal{C}'_{i_0-1}, \mathcal{C}'_{i_0+1}, \cdots, \mathcal{C}'_p\rangle = \mathcal{G}\langle(\eta'_{ij})_{1\leq i\leq p,\, i\neq i_0,\, 1\leq j\leq n}\rangle,$$

so that (Chapter I, Proposition 1) $\mathcal{G}_{i_0}\cdot\Lambda_{i_0}$ is prime $(1 \leq i_0 \leq p)$. It follows (see remark following the proof of Proposition 4 of Chapter I) that the ideal Λ generated by $\bigcup_{i=1}^{p}\Lambda_i$ in $\mathcal{G}\{(y_{ij})_{1\leq i\leq p,1\leq j\leq n}\}$ is a prime differential ideal. Since the order of each Λ_i is finite so is the order of Λ, and $\operatorname{ord}\Lambda = \sum \operatorname{ord}\Lambda_i$. Therefore (Chapter I, Corollary to Proposition 2) $(\sigma_i\eta_j)_{1\leq i\leq p,\,1\leq j\leq n}$ is a generic zero of Λ if and only if

$$\begin{aligned}\partial^0\mathcal{G}\langle\sigma_1\mathcal{G}, \cdots, \sigma_p\mathcal{G}\rangle/\mathcal{G} &= \partial^0\mathcal{G}\langle(\sigma_i\eta_j)_{1\leq i\leq p,\,1\leq j\leq n}\rangle/\mathcal{G} = \operatorname{ord}\Lambda\\ &= \sum\operatorname{ord}\Lambda_i = \sum\partial^0\mathcal{G}\langle\sigma_i\eta_1, \cdots, \sigma_i\eta_n\rangle/\mathcal{G} = \partial^0\mathcal{G}\langle\sigma_i\mathcal{G}\rangle/\mathcal{G}.\end{aligned}$$

It is easy to see, however, that $\sigma_1, \cdots, \sigma_p$ are independent if and only if $(\sigma_i\eta_j)_{1\leq i\leq p,1\leq j\leq n}$ is a generic zero of Λ. This completes the proof.

PROPOSITION 8. *If $\sigma_1, \cdots, \sigma_p$ are independent strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ and if $\tau_1, \cdots, \tau_p$ are isomorphisms of $\mathcal{G}$ such that τ_i is a specialization of σ_i $(1 \leq i \leq p)$ then $(\tau_1, \tau_1\tau_2, \cdots, \tau_1\tau_p)$ is a specialization of $(\sigma_1, \sigma_1\sigma_2, \cdots, \sigma_1\sigma_p)$.*

Proof. By (1) and (2) we have

$$\begin{aligned}\partial^0\mathcal{G}\langle\sigma_1^{-1}\mathcal{G}, \sigma_2\mathcal{G}, \cdots, \sigma_p\mathcal{G}\rangle/\mathcal{G} &= \partial^0\mathcal{G}\langle\mathcal{C}_{\sigma_1}, \mathcal{C}_{\sigma_2}, \cdots, \mathcal{C}_{\sigma_p}\rangle\mathcal{G}\\ &= \partial^0\mathcal{G}\langle\sigma_1\mathcal{G}, \sigma_2\mathcal{G}, \cdots, \sigma_p\mathcal{G}\rangle/\mathcal{G},\end{aligned}$$

which (Proposition 7) equals

$$\textstyle\sum_{1\leq i\leq p}\partial^0\mathcal{G}\langle\sigma_i\mathcal{G}\rangle/\mathcal{G} = \partial^0\mathcal{G}\langle\sigma_1^{-1}\mathcal{G}\rangle/\mathcal{G} + \sum_{2\leq i\leq p}\partial^0\mathcal{G}\langle\sigma_i\mathcal{G}\rangle/\mathcal{G}.$$

Therefore (Proposition 7) $\sigma_1^{-1}, \sigma_2, \cdots, \sigma_p$ are independent. If τ_i is a specialization of σ_i $(1 \leq i \leq p)$ then (Proposition 6) τ_1^{-1} is a specialization of σ_1^{-1}, whence $(\tau_1^{-1}, \tau_2, \cdots, \tau_p)$ is a specialization of $(\sigma_1^{-1}, \sigma_2, \cdots, \sigma_p)$. By Proposition 6 it follows that $(\tau_1, \tau_1\tau_2, \cdots, \tau_1\tau_p)$ is a specialization of $(\sigma_1, \sigma_1\sigma_2, \cdots, \sigma_1\sigma_p)$.

If $W(X_1, \cdots, X_p)$ is a word in $X_1, \cdots, X_p$, that is if it is an element of the free group generated by $X_1, \cdots, X_p$, and if $\sigma_1, \cdots, \sigma_p$ are strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, then $W(\sigma_1, \cdots, \sigma_p)$, the meaning of which is obvious, is itself a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$.

If σ is a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$ then, since $\mathcal{G}$ is finitely generated over $\mathcal{F}$, $\sigma\mathcal{G}$ is too, so that $\mathcal{G}\langle\sigma\mathcal{G}\rangle = \mathcal{G}\langle\mathcal{C}_\sigma\rangle$ is finitely generated over $\mathcal{G}$; using Chapter I, § 4, it is not difficult to see that then $\mathcal{C}_\sigma$ is finitely generated over $\mathcal{C}$.

PROPOSITION 9. *Let $\sigma_1, \cdots, \sigma_p$ be strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, let $\gamma_{i1}, \cdots, \gamma_{iq_i}$ be constants such that $\mathcal{C}_{\sigma_i} = \mathcal{C}\langle\gamma_{i1}, \cdots, \gamma_{iq_i}\rangle$ $(1 \leq i \leq p)$, let $W_1(X_1, \cdots, X_p), \cdots, W_r(X_1, \cdots, X_p)$ be words, let $\xi_1, \cdots, \xi_r$ be elements of $\mathcal{G}$, and let N be a differential polynomial in $\mathcal{G}\{w_1, \cdots, w_r\}$ which does not vanish at $(W_1(\sigma_1, \cdots, \sigma_p)\xi_1, \cdots, W_r(\sigma_1, \cdots, \sigma_p)\xi_r)$. Then there exists a polynomial $M \in \mathcal{C}[(u_{ik})_{1\leq i\leq p, 1\leq k\leq q_i}]$ which does not vanish at $(\gamma_{ik})_{1\leq i\leq p, 1\leq k\leq q_i}$ and which has the following property: if $(c_{ik})_{1\leq i\leq p, 1\leq k\leq q_i}$ is a family of constants which is a specialization of $(\gamma_{ik})_{1\leq i\leq p, 1\leq k\leq q_i}$ over $\mathcal{C}$ which does not annul M then for each i $(1 \leq i \leq p)$ there exists a unique isomorphism τ_i of $\mathcal{G}$ such that $((\tau_i\alpha)_{\alpha\in G}, (c_{ik})_{1\leq k\leq q_i})$ is a specialization of $((\sigma_i\alpha)_{\alpha\in G}, (\gamma_{ik})_{1\leq k\leq q_i})$ over $\mathcal{G}$, $\mathcal{C}_{\tau_i} = \mathcal{C}\langle c_{i1}, \cdots, c_{iq_i}\rangle$, N does not vanish at*

$$(W_1(\tau_1, \cdots, \tau_p)\xi_1, \cdots, W_r(\tau_1, \cdots, \tau_p)\xi_r),$$

and, for every finite family

$$(W'_1(X_1, \cdots, X_p), \cdots, W'_s(X_1, \cdots, X_p))$$

of words,

$$(W'_1(\tau_1, \cdots, \tau_p), \cdots, W'_s(\tau_1, \cdots, \tau_p))$$

is a specialization of

$$(W'_1(\sigma_1, \cdots, \sigma_p), \cdots, W'_s(\sigma_1, \cdots, \sigma_p)).$$

Proof. By (1) there exist

$$A_{ij}, B_{ij} \in \mathcal{F}\{y_1, \cdots, y_n\}[u_{i1}, \cdots, u_{iq_i}] \qquad (1 \leq i \leq p, 1 \leq j \leq n),$$
$$C_{ik}, D_{ik} \in \mathcal{F}\{y_1, \cdots, y_n, z_{i1}, \cdots, z_{in}\} \qquad (1 \leq i \leq p, 1 \leq k \leq q_i),$$
$$E_{ij}, F_{ij} \in \mathcal{F}\{z_{i1}, \cdots, z_{in}\}[u_{i1}, \cdots, u_{iq_i}] \qquad (1 \leq i \leq p, 1 \leq j \leq n)$$

such that, for all i, j, k,

$$B_{ij}(\eta_1, \cdots, \eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}) \neq 0,$$

$$D_{ik}(\eta_1, \cdots, \eta_n, \sigma_i\eta_1, \cdots, \sigma_i\eta_n) \neq 0,$$

$$F_{ij}(\sigma_i\eta_1, \cdots, \sigma_i\eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}) \neq 0,$$

$$(3) \qquad \sigma_i\eta_j = A_{ij}(\eta_1, \cdots, \eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i})/B_{ij}(\eta_1, \cdots, \eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}),$$

$$\gamma_{ik} = C_{ik}(\eta_1, \cdots, \eta_n, \sigma_i\eta_1, \cdots, \sigma_i\eta_n)/D_{ik}(\eta_1, \cdots, \eta_n, \sigma_i\eta_1, \cdots, \sigma_i\eta_n),$$

$$(4) \qquad \eta_j = E_{ij}(\sigma_i\eta_1, \cdots, \sigma_i\eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i})/F_{ij}(\sigma_i\eta_1, \cdots, \sigma_i\eta_n, \gamma_{i1}, \cdots, \gamma_{iq_i}).$$

Let $(c_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}$ be a family of constants which is a specialization of $(\gamma_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}$ over $\mathcal{C}$, and therefore (Chapter I, Corollary 2 to Proposition 3) over $\mathcal{G}$. If

$$(5) \qquad \prod_{i,j} B_{ij}(\eta_1, \cdots, \eta_n, c_{i1}, \cdots, c_{iq_i}) \neq 0$$

and if we set

$$(6) \qquad \zeta_{ij} = A_{ij}(\eta_1, \cdots, \eta_n, c_{i1}, \cdots, c_{iq_i})/B_{ij}(\eta_1, \cdots, \eta_n, c_{i1}, \cdots, c_{iq_i})$$

then $((\zeta_{ij})_{1\leq i\leq p,\, 1\leq j\leq n}, (c_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i})$ is a specialization of

$$((\sigma_i\eta_j)_{1\leq i\leq p,\, 1\leq j\leq n}, (\gamma_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i})$$

over $\mathcal{G}$. If moreover

$$(7) \qquad \prod_{i,k} D_{ik}(\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}) \neq 0$$

then

$$(8) \qquad c_{ik} = C_{ik}(\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in})/D_{ik}(\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}).$$

If in addition

$$(9) \qquad \prod_{i,j} F_{ij}(\zeta_{i1}, \cdots, \zeta_{in}, c_{i1}, \cdots, c_{iq_i}) \neq 0$$

then

$$(10) \qquad \eta_j = E_{ij}(\zeta_{i1}, \cdots, \zeta_{in}, c_{i1}, \cdots, c_{iq_i})/F_{ij}(\zeta_{i1}, \cdots, \zeta_{in}, c_{i1}, \cdots, c_{iq_i}).$$

Assuming that (5), (7), (9) hold we see that $(\zeta_{i1}, \cdots, \zeta_{in})$ is a specialization of $(\sigma_i\eta_1, \cdots, \sigma_i\eta_n)$ over $\mathcal{G}$, and therefore of $(\eta_1, \cdots, \eta_n)$ over $\mathcal{F}$, such that

$$\begin{aligned}
&\partial^0\mathcal{F}\langle\zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathcal{F} - \partial^0\mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle/\mathcal{F} \\
&= \partial^0\mathcal{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathcal{F} - \partial^0\mathcal{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathcal{F}\langle\zeta_{i1}, \cdots, \zeta_{in}\rangle \\
&\qquad - \partial^0\mathcal{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathcal{F} \\
&\qquad + \partial^0\mathcal{F}\langle\eta_1, \cdots, \eta_n, \zeta_{i1}, \cdots, \zeta_{in}\rangle/\mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle,
\end{aligned}$$

which by (6), (8), (10) equals

$$-\partial^0\mathcal{F}\langle\zeta_{i1},\cdots,\zeta_{in},c_{i1},\cdots,c_{iq_i}\rangle/\mathcal{F}\langle\zeta_{i1},\cdots,\zeta_{in}\rangle$$
$$+\,\partial^0\mathcal{F}\langle\eta_1,\cdots,\eta_n,c_{i1},\cdots,c_{iq_i}\rangle/\mathcal{F}\langle\eta_1,\cdots,\eta_n\rangle,$$

which by Chapter I, Corollary 3 to Proposition 3, is ≥ 0. It follows (Chapter I, Proposition 2) that $(\zeta_{i1},\cdots,\zeta_{in})$ is a generic specialization of $(\eta_1,\cdots,\eta_n)$ over $\mathcal{F}$, so that there exists a unique isomorphism τ_i of $\mathcal{G}=\mathcal{F}\langle\eta_1,\cdots,\eta_n\rangle$ over $\mathcal{F}$ such that $\tau_i\eta_j=\zeta_{ij}$ $(1\leq j\leq n)$. By (6), (8), (10)

$$\mathcal{G}\langle\tau_i\mathcal{G}\rangle=\mathcal{G}\langle c_{i1},\cdots,c_{iq_i}\rangle=(\tau_i\mathcal{G})\langle c_{i1},\cdots,c_{iq_i}\rangle,$$

so that $\mathcal{C}_{\tau_i}=\mathcal{C}\langle c_{i1},\cdots,c_{iq_i}\rangle$. It is apparent that τ_i is the unique isomorphism of $\mathcal{G}$ such that $((\tau_i\alpha)_{\alpha\varepsilon G},(c_{ik})_{1\leq k\leq q_i})$ is a specialization of $((\sigma_i\alpha)_{\alpha\varepsilon G},(\gamma_{ik})_{1\leq k\leq q_i})$ over $\mathcal{G}$.

By (4) and (10)

$$\sigma_i^{-1}\eta_j=E_{ij}(\eta_1,\cdots,\eta_n,\gamma_{i1},\cdots,\gamma_{iq_i})/F_{ij}(\eta_1,\cdots,\eta_n,\gamma_{i1},\cdots,\gamma_{iq_i}),\tag{11}$$

$$\tau_i^{-1}\eta_j=E_{ij}(\eta_1,\cdots,\eta_n,c_{i1},\cdots,c_{iq_i})/F_{ij}(\eta_1,\cdots,\eta_n,c_{i1},\cdots,c_{iq_i}).\tag{12}$$

From (3), (6), (11), (12) it is not difficult to see that for every word $W(X_1,\cdots,X_p)$ there exist

$$G_j^W,H_j^W\,\varepsilon\,\mathcal{F}\{y_1,\cdots,y_n\}[(u_{ik})_{1\leq i\leq p,\,1\leq k\leq q_i}]$$

such that

$$H_j^W(\eta_1,\cdots,\eta_n,(\gamma_{ik}))\neq 0,\qquad H_j^W(\eta_1,\cdots,\eta_n,(c_{ik}))\neq 0$$

and

$$W(\sigma_1,\cdots,\sigma_p)\eta_j=G_j^W(\eta_1,\cdots,\eta_n,(\gamma_{ik}))/H_j^W(\eta_1,\cdots,\eta_n,(\gamma_{ik})),$$
$$W(\tau_1,\cdots,\tau_p)\eta_j=G_j^W(\eta_1,\cdots,\eta_n,(c_{ik}))/H_j^W(\eta_1,\cdots,\eta_n,(c_{ik})).$$

It follows, for any finite family of words

$$W'_1(X_1,\cdots,X_p),\cdots,W'_s(X_1,\cdots,X_p),$$

that $(W'_l(\tau_1,\cdots,\tau_p)\eta_j)_{1\leq l\leq s,\,1\leq j\leq n}$ is a specialization of

$$W'_l(\sigma_1,\cdots,\sigma_p)\eta_j)_{1\leq l\leq s,\,1\leq j\leq n}$$

over $\mathcal{G}$, so that

$$(W'_1(\tau_1,\cdots,\tau_p),\cdots,W'_s(\tau_1,\cdots,\tau_p))$$

is a specialization of

$$(W'_1(\sigma_1,\cdots,\sigma_p),\cdots,W'_s(\sigma_1,\cdots,\sigma_p)).$$

It also follows that for each h $(1\leq h\leq r)$ there exist

$$I_h,J_h\,\varepsilon\,\mathcal{F}\{y_1,\cdots,y_n\}[(u_{ik})_{1\leq i\leq p,\,1\leq k\leq q_i}]$$

with

$$J_h(\eta_1, \cdots, \eta_n, (\gamma_{ik})) \neq 0, \qquad J_h(\eta_1, \cdots, \eta_n, (c_{ik})) \neq 0$$

such that

$$W_h(\sigma_1, \cdots, \sigma_p)\xi_h = I_h(\eta_1, \cdots, \eta_n, (\gamma_{ik}))/J_h(\eta_1, \cdots, \eta_n, (\gamma_{ik})),$$

$$W_h(\tau_1, \cdots, \tau_p)\xi_h = I_h(\eta_1, \cdots, \eta_n, (c_{ik}))/J_h(\eta_1, \cdots, \eta_n, (c_{ik})),$$

and therefore that there exist

$$U, V \in \mathcal{F}\{y_1, \cdots, y_n\}[(u_{ik})_{1 \leq i \leq p, 1 \leq k \leq q_i}]$$

such that

$$V(\eta_1, \cdots, \eta_n), (\gamma_{ik})) \neq 0, \qquad V(\eta_1, \cdots, \eta_n (c_{ik})) \neq 0$$

and

$$N(W_1(\sigma_1, \cdots, \sigma_p)\xi_1, \cdots, W_r(\sigma_1, \cdots, \sigma_p)\xi_r)$$
$$= U(\eta_1, \cdots, \eta_n, (\gamma_{ik}))/V(\eta_1, \cdots, \eta_n, (\gamma_{ik})),$$

$$N(W_1(\tau_1, \cdots, \tau_p)\xi_1, \cdots, W_r(\tau_1, \cdots, \tau_p)\xi_r)$$
$$= U(\eta_1, \cdots, \eta_n, (c_{ik}))/V(\eta_1, \cdots, \eta_n, (c_{ik})),$$

Now there exists a polynomial $M' \in \mathcal{G}[(u_{ik})_{1 \leq i \leq p, 1 \leq k \leq q_i}]$ which has the two properties that it does not vanish at $(\gamma_{ik})_{1 \leq i \leq p, 1 \leq k \leq q_i}$ and that if it does not vanish at $(c_{ik})_{1 \leq i \leq p, 1 \leq k \leq q_i}$ then (5), (7), (9) hold and $U(\eta_1, \cdots, \eta_n, (c_{ik})) \neq 0$. Therefore (Chapter I, Corollary 2 to Proposition 3) there exists a polynomial $M \in \mathcal{C}[(u_{ik})_{1 \leq i \leq p, 1 \leq k \leq q_i}]$ with the same two properties. This M has the property described in the statement of the proposition.

COROLLARY 1. *Let $\eta_1, \cdots, \eta_n$ be elements of $\mathcal{G}$ such that $\mathcal{G} = \mathcal{F}\langle \eta_1, \cdots, \eta_n \rangle$, let q, r be integers such that $1 \leq q \leq r$, let $W_1(X_1, \cdots, X_p), \cdots, W_r(X_1, \cdots, X_p)$ be words, let $\sigma_1, \cdots, \sigma_p$ be strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$, and let $Q \in \mathcal{G}\{(y_{ij})_{1 \leq i \leq r, 1 \leq j \leq n}\}$ not vanish at $(W_i(\sigma_1, \cdots, \sigma_p)\eta_j)_{1 \leq i \leq r, 1 \leq j \leq n}$. Then there exists a $P \in \mathcal{G}\{(y_{ij})_{1 \leq i \leq q, 1 \leq j \leq n}\}$ which does not vanish at $(W_i(\sigma_1, \cdots, \sigma_p)\eta_j)_{1 \leq i \leq q, 1 \leq j \leq n}$ such that for every specialization $(\rho'_1, \cdots, \rho'_q)$ of $(W_1(\sigma_1, \cdots, \sigma_p), \cdots, W_q(\sigma_1, \cdots, \sigma_p))$ for which $P((\rho'_i\eta_j)_{1 \leq i \leq q, 1 \leq j \leq n}) \neq 0$ there exists strong isomorphisms $\sigma'_1, \cdots, \sigma'_p$ of G such that $\rho'_i = W_i(\sigma'_1, \cdots, \sigma'_p)$ $(1 \leq i \leq q)$,*

$$(\sigma'_1, \cdots, \sigma'_p, W_1(\sigma'_1, \cdots, \sigma'_p), \cdots, W_r(\sigma'_1, \cdots, \sigma'_p))$$

is a specialization of

$$(\sigma_1, \cdots, \sigma_p, W_1(\sigma_1, \cdots, \sigma_p), \cdots, W_r(\sigma_1, \cdots, \sigma_p)),$$

and $Q((W_i(\sigma'_1, \cdots, \sigma'_p)\eta_j)_{1 \leq i \leq r, 1 \leq j \leq n}) \neq 0$.

COROLLARY 2. *Let σ be a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$, let $\xi_1, \cdots, \xi_s \in \mathcal{G}$, let $N \in \mathcal{G}\{w_1, \cdots, w_s\}$ not vanish at $(\sigma\xi_1, \cdots, \xi\sigma_s)$. Then there exists an automorphism τ of $\mathcal{G}$ over $\mathcal{F}$ such that τ is a specialization of σ and $N(\tau\xi_1, \cdots, \tau\xi_s) \neq 0$.*

COROLLARY 3. *Let σ be a strong isomorphism of $\mathcal{G}$ over $\mathcal{F}$, let $\mathcal{F}_1$ be a differential field between $\mathcal{F}$ and $\mathcal{G}$, and suppose that the restriction σ' of σ to $\mathcal{F}_1$ is a strong isomorphism of $\mathcal{F}_1$. Then there exist a finite number of nongeneric specializations $\sigma'_1, \cdots, \sigma'_s$ of σ' such that every specialization τ' of σ' which is not a specialization of σ'_j $(1 \leq j \leq s)$ is the restriction to $\mathcal{F}_1$ of some specialization τ of σ.*

Proof. Let $\gamma_1, \cdots, \gamma_p, \gamma_{p+1}, \cdots, \gamma_q$ be constants such that $\mathcal{F}_1\langle\sigma'\mathcal{F}_1\rangle = \mathcal{F}_1\langle\gamma_1, \cdots, \gamma_p\rangle$, $\mathcal{G}\langle\sigma\mathcal{G}\rangle = \mathcal{G}\langle\gamma_1, \cdots, \gamma_q\rangle$. By Proposition 9 there exists a polynomial $M \in \mathcal{C}[u_1, \cdots, u_q]$ with $M(\gamma_1, \cdots, \gamma_q) \neq 0$ such that whenever $c_1, \cdots, c_q$ are constants such that $(c_1, \cdots, c_q)$ is a specialization of $(\gamma_1, \cdots, \gamma_q)$ over $\mathcal{C}$ with $M(c_1, \cdots, c_n) \neq 0$ then there exist an isomorphism τ of $\mathcal{G}$ such that $((\tau\alpha)_{\alpha\in G}, c_1, \cdots, c_q)$ is a specialization of $((\sigma\alpha)_{\alpha\in G}, \gamma_1, \cdots, \gamma_q)$ over $\mathcal{G}$, and a unique isomorphism τ' of $\mathcal{F}_1$ such that $((\tau'\alpha)_{\alpha\in\mathcal{F}_1}, c_1, \cdots, c_p)$ is a specialization of $((\sigma'\alpha)_{\alpha\in\mathcal{F}_1}, \gamma_1, \cdots, \gamma_p)$ over $\mathcal{F}_1$. There also exists a polynomial $K \in \mathcal{C}[u_1, \cdots, u_p]$ with $K(\gamma_1, \cdots, \gamma_p) \neq 0$ such that every specialization $(c_1, \cdots, c_p)$ of $(\gamma_1, \cdots, \gamma_p)$ with $K(c_1, \cdots, c_p) \neq 0$ can be extended to a specialization $(c_1, \cdots, c_q)$ of $(\gamma_1, \cdots, \gamma_q)$ over $\mathcal{C}$ with $M(c_1, \cdots, c_q) \neq 0$. We now write, for $1 \leq i \leq p$,

$$\gamma_i = R_i(\sigma'\theta_1, \cdots, \sigma'\theta_l)/S(\sigma'\theta_1, \cdots, \sigma'\theta_l),$$

where $\theta_1, \cdots, \theta_l$ are elements of $\mathcal{F}_1$, such that $\mathcal{F}_1 = \mathcal{F}\langle\theta_1, \cdots, \theta_l\rangle$, R_i and $S \in \mathcal{F}_1\{y_1, \cdots, y_l\}$, and $S(\sigma'\theta_1, \cdots, \sigma'\theta_l) \neq 0$. If τ' is any specialization of σ' such that $S(\tau'\theta_1, \cdots, \tau'\theta_l) \neq 0$ and if we set

$$c_i = R_i(\tau'\theta_1, \cdots, \tau'\theta_l)/S(\tau'\theta_1, \cdots, \tau'\theta_l),$$

then $c_1, \cdots, c_p$ are constants such that $(c_1, \cdots, c_p)$ is a specialization of $(\gamma_1, \cdots, \gamma_p)$ over $\mathcal{C}$. If in addition

$$K\left(\frac{R_1(\tau'\theta_1, \cdots, \tau'\theta_l)}{S(\tau'\theta_1, \cdots, \tau'\theta_l)}, \cdots, \frac{R_p(\tau'\theta_1, \cdots, \tau'\theta_l)}{S(\tau'\theta_1, \cdots, \tau'\theta_l)}\right) \neq 0,$$

that is $K(c_1, \cdots, c_p) \neq 0$, then we may extend $(c_1, \cdots, c_p)$ to a specialization $(c_1, \cdots, c_q)$ of $(\gamma_1, \cdots, \gamma_q)$ over $\mathcal{C}$ such that $M(c_1, \cdots, c_q) \neq 0$, so that there exists an isomorphism τ of $\mathcal{G}$ such that $((\tau\alpha)_{\alpha\in G}, c_1, \cdots, c_q)$ is a specialization of $((\sigma\alpha)_{\alpha\in G}, \gamma_1, \cdots, \gamma_q)$ over $\mathcal{G}$. Under these circum-

stances τ will be a specialization of σ, and τ' will be the restriction of τ to $\mathcal{F}_1$. Thus, the specializations τ' of σ' which can *not* be extended to a specialization of σ have the property that $(\tau'\theta_1, \cdots, \tau'\theta_l)$ is a zero of SL, where L is a differential polynomial in $\mathcal{F}_1\{y_1, \cdots, y_l\}$ obtained by multiplying $K(S^{-1}R_1, \cdots, S^{-1}R_l)$ by a power of S; of course $(\sigma'\theta_1, \cdots, \sigma'\theta_l)$ is not a zero of SL. Now let $\mathfrak{S}$ be the prime differential ideal of $\mathcal{F}_1\{y_1, \cdots, y_l\}$ with generic zero $(\sigma'\theta_1, \cdots, \sigma'\theta_l)$, and let $\mathfrak{S}_1, \cdots, \mathfrak{S}_s$ be the minimal prime differential ideal divisors of $\{\mathfrak{S}, SL\}$ in $\mathcal{F}_1\{y_1, \cdots, y_l\}$. If τ' is a specialization of σ' which can not be extended to a specialization of σ then $(\tau'\theta_1, \cdots, \tau'\theta_l)$ is a zero of $\mathfrak{S}_j$ for some j. Let $(\psi_{j1}, \cdots, \psi_{jl})$ be a generic zero of this $\mathfrak{S}_j$; then $(\psi_{j1}, \cdots, \psi_{jl})$ is a zero of $\mathfrak{S}$. hence a specialization of $(\sigma'\theta_1, \cdots, \sigma'\theta_l)$ over $\mathcal{F}_1$ and a fortiori over $\mathcal{F}$, and therefore a specialization of $(\theta_1, \cdots, \theta_l)$ over $\mathcal{F}$. But $(\psi_{j1}, \cdots, \psi_{jl})$ admits $(\tau'\theta_1, \cdots, \tau'\theta_l)$ as a specialization over $\mathcal{F}_1$ and therefore over $\mathcal{F}$, so that $(\theta_1, \cdots, \theta_l)$ is a specialization of $(\psi_{j1}, \cdots, \psi_{jl})$ over $\mathcal{F}$. Thus $(\psi_{j1}, \cdots, \psi_{jl})$ is a generic specialization of $(\theta_1, \cdots, \theta_l)$ over $\mathcal{F}$, so that there is a unique isomorphism σ'_j of $\mathcal{F}_1$ over $\mathcal{F}$ such that $\sigma'_j\theta_i = \psi_{ji}$ $(1 \leq i \leq l)$. Clearly τ' is a specialization of σ'_j, σ'_j is a specialization of σ', and because

$$\partial^0\sigma'_j\mathcal{F}_1/\mathcal{F}_1 = \operatorname{ord}\mathfrak{S}_j < \operatorname{ord}\mathfrak{S} = \partial^0\sigma'\mathcal{F}_1/\mathcal{F}_1$$

the latter specialization is nongeneric. This completes the proof of Corollary 3.

5. Algebraic sets. A subset $\mathfrak{M}^*$ of the group $\mathfrak{G}^*$ of all strong isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ will be called an *irreducible set in* $\mathfrak{G}^*$ if $\mathfrak{M}^*$ contains an element σ^* such that $\mathfrak{M}^*$ is the set of all specializations of σ^*; any such σ^* will then be called a *generic element* of $\mathfrak{M}^*$, and the transcendence degree $\partial^0\mathcal{G}\langle\sigma^*\mathcal{G}\rangle/\mathcal{G}$, which does not depend on the particular generic element σ^* employed, will be called the *dimension* of $\mathfrak{M}^*$ (notation: $\dim\mathfrak{M}^*$). It follows from Chapter I, Projosition 2, that if σ^* and τ^* are any two isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ such that τ^* is a specialization of σ^* then $\partial^0\mathcal{G}\langle\tau^*\mathcal{G}\rangle/\mathcal{G} \leq \partial^0\mathcal{G}\langle\sigma^*\mathcal{G}\rangle/\mathcal{G}$, and we have equality here if and only if the specialization is generic. Consequently if $\mathfrak{M}^*$ and $\mathfrak{N}^*$ are irreducible sets in $\mathfrak{G}^*$ such that $\mathfrak{M}^* \supseteq \mathfrak{N}^*$ then $\dim\mathfrak{M}^* \geq \dim\mathfrak{N}^*$, and $\dim\mathfrak{M}^* = \dim\mathfrak{N}^*$ if and only if $\mathfrak{M}^* = \mathfrak{N}^*$. A subset $\mathfrak{M}^*$ of $\mathfrak{G}^*$ will be called an *algebraic set in* $\mathfrak{G}^*$ if $\mathfrak{M}^*$ is the union of a finite number of irreducible sets in $\mathfrak{G}^*$. It is easy to see that an algebraic set in $\mathfrak{G}^*$ can be written in one and only one way as the union of a finite set of irreducible sets in $\mathfrak{G}^*$ none of which contains another; these unique irreducible sets, which are the maximal irreducible sets contained in the algebraic set, will be called the *components* of the algebraic set.

It is a consequence of Corollary 2 of Proposition 9 that every nonempty algebraic set in $\mathfrak{G}^*$ has a nonempty intersection with the group $\mathfrak{G}$ of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$, and that if $\mathfrak{M}^*$, $\mathfrak{N}^*$ are algebraic sets in $\mathfrak{G}^*$ with $\mathfrak{M}^* \neq \mathfrak{N}^*$ then $\mathfrak{M}^* \cap \mathfrak{G} \neq \mathfrak{N}^* \cap \mathfrak{G}$. Accordingly we shall call a subset $\mathfrak{M}$ of $\mathfrak{G}$ an *irreducible set in* $\mathfrak{G}$ or an *algebraic set in* $\mathfrak{G}$ if $\mathfrak{M}$ is the intersection with $\mathfrak{G}$ of, respectively, an irreducible set in $\mathfrak{G}^*$ or an algebraic set in $\mathfrak{G}^*$; this set in $\mathfrak{G}^*$, which is unique, we shall denote by $\mathfrak{M}^*$. If $\mathfrak{M}$ is an irreducible set in $\mathfrak{G}$ we define the *dimension* of $\mathfrak{M}$ (notation: $\dim \mathfrak{M}$) by the formula $\dim \mathfrak{M} = \dim \mathfrak{M}^*$, and define *generic element* of $\mathfrak{M}$ as a generic element of $\mathfrak{M}^*$ (so that a generic element of $\mathfrak{M}$ need not be an element of $\mathfrak{M}$); if σ^* is a generic element of $\mathfrak{M}$ we thus have $\dim \mathfrak{M} = \partial^0 \mathcal{G}\langle\sigma^* \mathcal{G}\rangle/\mathcal{G}$. If $\mathfrak{M}$ and $\mathfrak{N}$ are two irreducible sets in $\mathfrak{G}$ with $\mathfrak{M} \supseteq \mathfrak{N}$ then $\dim \mathfrak{M} \geq \dim \mathfrak{N}$, and equality of dimension implies that $\mathfrak{M} = \mathfrak{N}$. An algebraic set in $\mathfrak{G}$ can be written in one and only one way as the union of a finite set of irreducible sets in $\mathfrak{G}$ none of which contains another; these unique irreducible sets, which are the maximal irreducible sets in $\mathfrak{G}$ contained in the algebraic set, will be called the *components* of the algebraic set. If $\mathfrak{M}$ is an algebraic set in $\mathfrak{G}$ and $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$ are its components then $\mathfrak{M}^*_1, \cdots, \mathfrak{M}^*_p$ are the components of $\mathfrak{M}^*$.

PROPOSITION 10. *Every nonempty set of algebraic sets in* $\mathfrak{G}$ *has a minimal element.*

Proof. To each algebraic set $\mathfrak{M}$ we associate the sequence

$$k(\mathfrak{M}) = (k_d(\mathfrak{M}))_{0 \leq d < \infty},$$

where $k_d(\mathfrak{M})$ is the number of components of $\mathfrak{M}$ of dimension d; since the dimension of every irreducible set is $\leq \partial^0 \mathcal{G}/\mathcal{F}$ we have $k_d(\mathfrak{M}) = 0$ for all $d > \partial^0 \mathcal{G}/\mathcal{F}$. We introduce an order into the set of sequences $k(\mathfrak{M})$ by writing $k(\mathfrak{M}) \leq k(\mathfrak{N})$ whenever either $k(\mathfrak{M}) = k(\mathfrak{N})$ or $k(\mathfrak{M}) \neq k(\mathfrak{N})$ and the last nonzero difference $k_d(\mathfrak{N}) - k_d(\mathfrak{M})$ is positive. It is easy to see that if $\mathfrak{M} \subseteq \mathfrak{N}$ then $k(\mathfrak{M}) \leq k(\mathfrak{N})$; since it is obvious that in any nonempty set of algebraic sets in G there exists one for which the associated sequence is minimal, the proof is complete.

PROPOSITION 11. *If* $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$ *are irreducible sets in* $\mathfrak{G}$ *then there exists an independent family of generic elements of* $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$; *if* $\rho_i \in \mathfrak{M}^*_i$ $(1 \leq i \leq p)$ *then* $\rho_1, \cdots, \rho_p$ *are independent generic elements of* $\mathfrak{M}_1, \cdots, \mathfrak{M}_p$ *if and only if* $\partial^0 \mathcal{G}\langle\rho_1 \mathcal{G}, \cdots, \rho_p \mathcal{G}\rangle/\mathcal{G} = \sum_{i=1}^{p} \dim \mathfrak{M}_i$.

Proof. Let σ_i be a generic element of $\mathfrak{M}_i$, and let $\gamma_{i1}, \cdots, \gamma_{iq_i}$ be con-

stants such that $\mathcal{B}_{\sigma_i} = \mathcal{B}\langle\gamma_{i1}, \cdots, \gamma_{iq_i}\rangle$, $1 \leq i \leq p$. It is obvious that there exist generic specializations $(\delta_{i1}, \cdots, \delta_{iq_i})$ of $(\gamma_{i1}, \cdots, \gamma_{iq_i})$ over $\mathcal{B}$, $1 \leq i \leq p$, such that

$$\partial^0 \mathcal{B}\langle(\delta_{ij})_{1\leq i\leq p,\, 1\leq j\leq q_i}\rangle/\mathcal{B} = \sum_{i=1}^{p} \partial^0 \mathcal{B}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle/\mathcal{B}.$$

By Proposition 9 there exists a strong isomorphism τ_i of $\mathcal{G}$ over $\mathcal{F}$ such that τ_i is a specialization of σ_i and $\mathcal{B}_{\tau_i} = \mathcal{B}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle$. Because

$$\begin{aligned}\partial^0 \mathcal{G}\langle\tau_i\mathcal{G}\rangle/\mathcal{G} &= \partial^0 \mathcal{B}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle/\mathcal{B} \\ &= \partial^0 \mathcal{B}\langle\gamma_{i1}, \cdots, \gamma_{iq_i}\rangle/\mathcal{B} = \partial^0 \mathcal{G}\langle\sigma_i\mathcal{G}\rangle/\mathcal{G}\end{aligned}$$

it follows that

$$\partial^0 \mathcal{G}\langle\tau_i\eta_1, \cdots, \tau_i\eta_n\rangle/\mathcal{G} = \partial^0 \mathcal{G}\langle\sigma_i\eta_1, \cdots, \sigma_i\eta_n\rangle/\mathcal{G},$$

where $\eta_1, \cdots, \eta_n$ are elements of $\mathcal{G}$ such that $\mathcal{G} = \mathcal{F}\langle\eta_1, \cdots, \eta_n\rangle$; therefore (Chapter I, Proposition 2) $(\tau_i\eta_1, \cdots, \tau_i\eta_n)$ is a generic specialization of $(\sigma_i\eta_1, \cdots, \sigma_i\eta_n)$ over $\mathcal{G}$, so that τ_i is a generic specialization of σ_i and therefore a generic element of $\mathfrak{M}_i$. Because

$$\begin{aligned}\partial^0 \mathcal{G}\langle\tau_1\mathcal{G}, \cdots, \tau_p\mathcal{G}\rangle/\mathcal{G} &= \partial^0 \mathcal{G}\langle(\delta_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}\rangle/\mathcal{G} \\ &= \partial^0 \mathcal{B}\langle(\delta_{ik})_{1\leq i\leq p,\, 1\leq k\leq q_i}\rangle/\mathcal{B} = \sum \partial^0 \mathcal{B}\langle\delta_{i1}, \cdots, \delta_{iq_i}\rangle/\mathcal{B} \\ &= \sum \partial^0 \mathcal{B}_{\tau_i}/\mathcal{B} = \sum \partial^0 \mathcal{G}\langle\tau_i\mathcal{G}\rangle/\mathcal{G},\end{aligned}$$

we see from Proposition 7 that $\tau_1, \cdots, \tau_p$ are independent. If ρ_i is an element of $\mathfrak{M}^*_i$ $(1 \leq i \leq p)$ then $\rho_1, \cdots, \rho_p$ are independent (Proposition 7) if and only if $\partial^0 \mathcal{G}\langle\rho_1\mathcal{G}, \cdots, \rho_p\mathcal{G}\rangle/\mathcal{G} = \sum \partial^0 \mathcal{G}\langle\rho_i\mathcal{G}\rangle/\mathcal{G}$, and are generic if and only if $\partial^0 \mathcal{G}\langle\rho_i\mathcal{G}\rangle/\mathcal{G} = \dim \mathfrak{M}_i$ $(1 \leq i \leq p)$.

PROPOSITION 12. *If $\mathfrak{M}$ is an irreducible set in $\mathfrak{G}$ with generic element σ and if $\tau \in \mathfrak{G}$ then $\tau\mathfrak{M}$ and $\mathfrak{M}\tau$ and $\mathfrak{M}^{-1}$ are irreducible sets in $\mathfrak{G}$ with generic elements $\tau\sigma$ and $\sigma\tau$ and σ^{-1} respectively, and* $\dim \tau\mathfrak{M} = \dim \mathfrak{M}\tau = \dim \mathfrak{M}^{-1} = \dim \mathfrak{M}$.

Proof. The set consisting of the single element τ is obviously an irreducible set of dimension 0, and

$$\partial^0 \mathcal{G}\langle\sigma\mathcal{G}, \tau\mathcal{G}\rangle/\mathcal{G} = \partial^0 \mathcal{G}\langle\sigma\mathcal{G}\rangle/\mathcal{G} = \dim \mathfrak{M} + 0;$$

therefore by Proposition 11 σ, τ are independent, so that (Proposition 8) every element of $\tau\mathfrak{M}$ is a specialization of $\tau\sigma$. Therefore if $\mathfrak{N}$ is the irreducible set in $\mathfrak{G}$ with generic element $\tau\sigma$ then $\tau\mathfrak{M} \subseteq \mathfrak{N}$. Similarly, $\tau^{-1}\mathfrak{N}$ is contained in the irreducible set in $\mathfrak{G}$ with generic element $\tau^{-1}(\tau\sigma) = \sigma$, that is in $\mathfrak{M}$, so that $\mathfrak{N} \subseteq \tau\mathfrak{M}$. Therefore $\tau\mathfrak{M} = \mathfrak{N}$, so that $\tau\mathfrak{M}$ is irreducible and

has generic element $\tau\sigma$. The proof for $\mathfrak{M}\tau$ is similar and for $\mathfrak{M}^{-1}$ is even simpler. Finally

$$\dim \tau\mathfrak{M} = \partial^0 \mathcal{G}\langle\tau\sigma\mathcal{G}\rangle/\mathcal{G} = \partial^0(\tau^{-1}\mathcal{G})\langle\sigma\mathcal{G}\rangle/\tau^{-1}\mathcal{G}$$
$$= \partial^0\mathcal{G}\langle\sigma\mathcal{G}\rangle/\mathcal{G} = \dim\mathfrak{M},$$

$$\dim\mathfrak{M}\tau = \partial^0\mathcal{G}\langle\sigma\tau\mathcal{G}\rangle/\mathcal{G} = \partial^0\mathcal{G}\langle\sigma\mathcal{G}\rangle/\mathcal{G} = \dim\mathfrak{M},$$

and $\dim\mathfrak{M}^{-1} = \dim\mathfrak{M}$ by (2).

PROPOSITION 13. *If Φ is a nonempty set of algebraic sets in $\mathfrak{G}$ then $\bigcap_{\mathfrak{M}\in\Phi}\mathfrak{M}$ is an algebraic set in $\mathfrak{G}$.*

Proof. If Φ contains only one element the conclusion is obvious. Suppose next that Φ contains precisely two elements, say $\mathfrak{M}_1$ and $\mathfrak{M}_2$, and assume first that $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are irreducible; let σ_1 and σ_2 be generic elements of $\mathfrak{M}_1$ and $\mathfrak{M}_2$ respectively. Letting $\eta_1,\cdots,\eta_n$ be elements of $\mathcal{G}$ such that $\mathcal{G} = \mathcal{F}\langle\eta_1,\cdots,\eta_n\rangle$, we let M_i denote the prime differential ideal in $\mathcal{G}\{y_1,\cdots,y_n\}$ with generic zero $(\sigma_i\eta_1,\cdots,\sigma_i\eta_n)$, $i=1,2$. Let $\mathrm{N}_1,\cdots,\mathrm{N}_p$ be the minimal prime differential ideal divisors of the perfect differential ideal $\{\mathrm{M}_1,\mathrm{M}_2\}$ in $\mathcal{G}\{y_1,\cdots,y_n\}$, and let $(\zeta_{i1},\cdots,\zeta_{in})$ be a generic zero of N_i, $1\leq i\leq p$. It is clear that $(\zeta_{i1},\cdots,\zeta_{in})$ is a specialization of $(\sigma_1\eta_1,\cdots,\sigma_1\eta_n)$ and of $(\sigma_2\eta_1,\cdots,\sigma_2\eta_n)$ over $\mathcal{G}$ and a fortiori over $\mathcal{F}$, so that $(\zeta_{i1},\cdots,\zeta_{in})$ is a specialization of $(\eta_1,\cdots,\eta_n)$ over $\mathcal{F}$. For every $\sigma\in\mathfrak{M}_1\cap\mathfrak{M}_2$, $(\sigma\eta_1,\cdots,\sigma\eta_n)$ is a zero of $\{\mathrm{M}_1,\mathrm{M}_2\}$ and hence of some N_i and therefore is a specialization of $(\zeta_{i1},\cdots,\zeta_{in})$ over $\mathcal{G}$ for some i. Let I be the set of all integers i with $1\leq i\leq p$ such that $\mathfrak{M}_1\cap\mathfrak{M}_2$ contains an element σ for which $(\sigma\eta_1,\cdots,\sigma\eta_n)$ is a specialization of $(\zeta_{i1},\cdots,\zeta_{in})$. For each $i\in I$ there exists a $\sigma\in\mathfrak{M}_1\cap\mathfrak{M}_2$ such that $(\sigma\eta_1,\cdots,\sigma\eta_n)$ is a specialization of $(\zeta_{i1},\cdots,\zeta_{in})$ over $\mathcal{G}$ and a fortiori over $\mathcal{F}$, so that $(\eta_1,\cdots,\eta_n)$ is a specialization of $(\zeta_{i1},\cdots,\zeta_{in})$ over $\mathcal{F}$. By the above, $(\zeta_{i1},\cdots,\zeta_{in})$ is then a generic specialization of $(\eta_1,\cdots,\eta_n)$ over $\mathcal{F}$, so that there exists a unique isomorphism τ_i of $\mathcal{G}$ over $\mathcal{F}$ such that $\tau_i\eta_1=\zeta_{i1},\cdots,\tau_i\eta_n=\zeta_{in}$. Let $\mathfrak{N}_i$ be the irreducible set in $\mathfrak{G}$ with generic element τ_i. Each element of $\mathfrak{M}_1\cap\mathfrak{M}_2$ is a specialization of some τ_i and therefore belongs to $\bigcup_{i\in I}\mathfrak{N}_i$. Conversely, each element of $\bigcup_{i\in I}\mathfrak{N}_i$ is a specialization of some τ_i and hence also of σ_1 and σ_2 and therefore belongs to $\mathfrak{M}_1\cap\mathfrak{M}_2$. Thus $\mathfrak{M}_1\cap\mathfrak{M}_2=\bigcup_{i\in I}\mathfrak{N}_i$, so that $\mathfrak{M}_1\cap\mathfrak{M}_2$ is algebraic.

Continuing with the case in which Φ consists of $\mathfrak{M}_1$ and $\mathfrak{M}_2$, now abandon the assumption that $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are irreducible. Denoting the components

of $\mathfrak{M}_i$ by $\mathfrak{M}_{i1}, \cdots, \mathfrak{M}_{ip_i}$ $(i = 1, 2)$, we have $\mathfrak{M}_1 \cap \mathfrak{M}_2 = \bigcup_{j,k} \mathfrak{M}_{1j} \cap \mathfrak{M}_{2k}$. By what has already been proved, each $\mathfrak{M}_{1j} \cap \mathfrak{M}_{2k}$ is algebraic, so that $\mathfrak{M}_1 \cap \mathfrak{M}_2$ is too. The proposition thus holds when Φ consists of two algebraic sets. The extension to the case in which Φ consists of any finite number of algebraic sets is immediate by induction.

Now let Φ be arbitrary. Let Ψ be the set of all intersections of finite nonempty subsets of Φ. By the case already known, each element of Ψ is an algebraic set in $\mathfrak{G}$. By Proposition 10, Ψ has a minimal element, say $\mathfrak{N}$. It is obvious that $\mathfrak{M} \cap \mathfrak{N} \in \Psi$ for each $\mathfrak{M} \in \Phi$; since $\mathfrak{M} \cap \mathfrak{N} \subseteq \mathfrak{N}$ and $\mathfrak{N}$ is minimal in Ψ this means that $\mathfrak{M} \cap \mathfrak{N} = \mathfrak{N}$, that is $\mathfrak{N} \subseteq \mathfrak{M}$. Therefore $\mathfrak{N} = \bigcap_{\mathfrak{M} \in \Phi} \mathfrak{M}$, and the latter is algebraic.

6. Algebraic groups. By an *algebraic group* in $\mathfrak{G}$ we shall mean a subset of $\mathfrak{G}$ which is a subgroup of $\mathfrak{G}$ and at the same time an algebraic set in $\mathfrak{G}$. If $\mathfrak{H}$ is an algebraic group in $\mathfrak{G}$ and $\mathfrak{H}^*$ is the algebraic set in $\mathfrak{G}^*$ such that $\mathfrak{H} = \mathfrak{G} \cap \mathfrak{H}^*$ then $\mathfrak{H}^*$ is a subgroup of $\mathfrak{G}^*$. To prove this we observe that for two independent generic points σ^*, τ^* of two not necessarily distinct components of $\mathfrak{H}^*$ (see Proposition 11) we have $\sigma^{*-1}\tau^* \in \mathfrak{H}^*$, for otherwise by Proposition 9 we could find elements $\sigma, \tau \in \mathfrak{H}$ such that $\sigma^{-1}\tau \notin \mathfrak{H}$; it follows from Proposition 6 that $\sigma^{-1}\tau \in \mathfrak{H}^*$ whenever $\sigma, \tau \in \mathfrak{H}^*$, so that $\mathfrak{H}^*$ is a group.

THEOREM 1. *Let $\mathfrak{H}$ be an algebraic group in $\mathfrak{G}$. The components of $\mathfrak{H}$ are pairwise disjoint; the component $\mathfrak{H}^0$ of $\mathfrak{H}$ which contains the identity automorphism ι is a normal algebraic subgroup of $\mathfrak{H}$ of finite index; the components of $\mathfrak{H}$ are the cosets of $\mathfrak{H}^0$ in $\mathfrak{H}$.*

Proof. Let $\mathfrak{H}_1$, $\mathfrak{H}_2$ be components of $\mathfrak{H}$ which contain ι and let σ^*_1, σ^*_2 be independent generic elements of $\mathfrak{H}_1$, $\mathfrak{H}_2$. Since $\sigma^*_1\sigma^*_2$ belongs to $\mathfrak{H}^*$ it belongs to some component $\mathfrak{H}^*_0$ of $\mathfrak{H}^*$. By Proposition 8, $\sigma^*_1 = \sigma^*_1\iota$ is a specialization of $\sigma^*_1\sigma^*_2$ so that $\mathfrak{H}^*_1 \subseteq \mathfrak{H}^*_0$ and therefore $\mathfrak{H}^*_1 = \mathfrak{H}^*_0$, and similarly $\sigma^*_2 = \iota\sigma^*_2$ is a specialization of $\sigma^*_1\sigma^*_2$ so that $\mathfrak{H}^*_2 \subseteq \mathfrak{H}^*_0$ and $\mathfrak{H}^*_2 = \mathfrak{H}^*_0$; therefore $\mathfrak{H}^*_1 = \mathfrak{H}^*_2$ so that $\mathfrak{H}_1 = \mathfrak{H}_2$. Thus precisely one of the components of $\mathfrak{H}$ contains ι; we denote this component by $\mathfrak{H}^0$. Now let $\mathfrak{H}'$ be any component of $\mathfrak{H}$; if σ' is any element of $\mathfrak{H}'$ then (Proposition 12) $\sigma'^{-1}\mathfrak{H}'$ is an irreducible subset of $\mathfrak{H}$ containing ι so that $\sigma'^{-1}\mathfrak{H}' \subseteq \mathfrak{H}^0$ and $\mathfrak{H}' \subseteq \sigma'\mathfrak{H}^0$. Since (Proposition 12) $\sigma'\mathfrak{H}^0$ is an irreducible subset of $\mathfrak{H}$ and $\mathfrak{H}'$ is a component of $\mathfrak{H}$ this implies that $\mathfrak{H}' = \sigma'\mathfrak{H}^0$. Similarly we find that $\mathfrak{H}' = \mathfrak{H}^0\sigma'$, so that $\sigma'\mathfrak{H}^0 = \mathfrak{H}^0\sigma'$. If we choose $\mathfrak{H}'$ as $\mathfrak{H}^0$ we see that $\sigma\mathfrak{H}^0 = \mathfrak{H}^0$

for every $\sigma \in \mathfrak{H}^0$ so that $\mathfrak{H}^0$ is a subgroup of $\mathfrak{H}$; since $\sigma'\mathfrak{H}^0 = \mathfrak{H}^0\sigma'$ for every $\sigma' \in \mathfrak{H}$, $\mathfrak{H}^0$ is a normal subgroup of $\mathfrak{H}$. By the above, the components of $\mathfrak{H}$ are the cosets of $\mathfrak{H}^0$ in $\mathfrak{H}$, and are therefore pairwise disjoint.

COROLLARY 1. *The components of an algebraic group in $\mathfrak{G}$ all have the same dimension.*

Proof. This follows from the Theorem and Proposition 12.

By the *dimension* of an algebraic group $\mathfrak{H}$ in $\mathfrak{G}$ we shall mean the dimension of any one of its components. The component containing ι we shall call the *component of the identity* of $\mathfrak{H}$, and we shall always denote it by $\mathfrak{H}^0$.

COROLLARY 2. *An algebraic group in $\mathfrak{G}$ is irreducible if and only if it has no algebraic subgroup of finite index > 1.*

Proof. If $\mathfrak{H}$ is not irreducible then $\mathfrak{H}^0$ is an algebraic subgroup of $\mathfrak{H}$ of finite index > 1. Conversely, if $\mathfrak{K}$ is an algebraic subgroup of $\mathfrak{H}$ of finite index > 1 then so is the component of the identity $\mathfrak{K}^0$ of $\mathfrak{K}$; as $\mathfrak{H}$ is the union of the left cosets of $\mathfrak{K}^0$ in $\mathfrak{H}$ and these are irreducible sets (Proposition 12) none of which contains another, $\mathfrak{H}$ is not irreducible.

PROPOSITION 14. *Let $\mathfrak{h}$ be a subgroup of $\mathfrak{G}$ which is contained in at least one algebraic set in $\mathfrak{G}$. Then the intersection $\mathfrak{H}$ of all the algebraic sets in $\mathfrak{G}$ which contain $\mathfrak{h}$ is an algebraic group in $\mathfrak{G}$.*

Proof. By Proposition 13, $\mathfrak{H}$ is an algebraic set in $\mathfrak{G}$. Obviously $\mathfrak{H}$ can be characterized as the smallest algebraic set in $\mathfrak{G}$ which contains $\mathfrak{h}$. If σ is any element of $\mathfrak{h}$ then $\mathfrak{H}\sigma$, which by Proposition 12 is an algebraic set, is obviously the smallest algebraic set containing $\mathfrak{h}\sigma$. Since $\mathfrak{h}\sigma = \mathfrak{h}$ this means that $\mathfrak{H}\sigma = \mathfrak{H}$, so that $\mathfrak{H}\mathfrak{h} = \mathfrak{H}$. Now let σ be any element of $\mathfrak{H}$. Obviously $\sigma\mathfrak{H}$ is the smallest algebraic set containing $\sigma\mathfrak{h}$; but $\sigma\mathfrak{h} \subseteq \mathfrak{H}\mathfrak{h} = \mathfrak{H}$, so that $\sigma\mathfrak{H} \subseteq \mathfrak{H}$. By Proposition 12, $\mathfrak{H}^{-1}$ is algebraic and therefore is the smallest algebraic set containing $\mathfrak{h}^{-1}$; since $\mathfrak{h}^{-1} = \mathfrak{h}$ this means that $\mathfrak{H}^{-1} = \mathfrak{H}$, so that $\mathfrak{H}$ is a group.

PROPOSITION 15. *Let $\mathfrak{H}$ be an algebraic group in $\mathfrak{G}$ and let $\mathfrak{h}$ be a subgroup of $\mathfrak{H}$ such that $\mathfrak{H} - \mathfrak{h}$ is contained in the union of a finite family of irreducible sets in $\mathfrak{G}$ each of dimension $< \dim \mathfrak{H}$. Then $\mathfrak{h} = \mathfrak{H}$.*

Proof. Expressing $\mathfrak{H}$ as the union of cosets of $\mathfrak{h}$, $\mathfrak{H} = \bigcup_{i \in I} \sigma_i\mathfrak{h}$ $(\sigma_{i_0} = \iota)$, we see that if there were more than one coset there would exist an $i_1 \in I$

with $i_0 \neq i_1$; for this i_1 we would have $\sigma_{i_1}\mathfrak{h} \subseteq \mathfrak{H} - \mathfrak{h}$, so that $\sigma_{i_1}\mathfrak{h}$ would be contained in a finite union of irreducible sets each of dimension $< \dim \mathfrak{H}$, whence (Proposition 12) so would $\mathfrak{h}$. This would imply that the same is true of $\mathfrak{H} = \mathfrak{h} \cup (\mathfrak{H} - \mathfrak{h})$, which is impossible.

PROPOSITION 16. *Let $\mathfrak{H}$ be an algebraic group in $\mathfrak{G}$, $\mathfrak{K}$ an algebraic subgroup of $\mathfrak{H}$, $\mathfrak{N}$ a normal algebraic subgroup of $\mathfrak{H}$. Then $\mathfrak{K}\mathfrak{N}$ is an algebraic group in $\mathfrak{G}$.*

Proof. $\mathfrak{K}\mathfrak{N}$ is a group; we must prove it is algebraic. Let (Proposition 11) $\kappa_1, \cdots, \kappa_r, \nu_1, \cdots, \nu_s$ be independent generic elements of the r components of $\mathfrak{K}$ and the s components of $\mathfrak{N}$, let $\mathfrak{M}_{ij}$ be the irreducible set in $\mathfrak{G}$ with generic element $\kappa_i\nu_j$, and let $\mathfrak{M}$ be the intersection of all the algebraic sets in $\mathfrak{G}$ which contain $\mathfrak{K}\mathfrak{N}$. By Proposition 8, $\mathfrak{K}\mathfrak{N} \subseteq \bigcup \mathfrak{M}_{ij}$, so that (Proposition 14) $\mathfrak{M}$ is an algebraic group and $\mathfrak{M} \subseteq \bigcup \mathfrak{M}_{ij}$. If we had $\kappa_i\nu_j \notin \mathfrak{M}^*$ then (Proposition 9) there would exist $\kappa' \varepsilon \mathfrak{K}$, $\nu'_j \varepsilon \mathfrak{N}$ such that $\kappa'_i\nu'_j \notin \mathfrak{M}$, which is impossible; it follows that $\mathfrak{M} = \bigcup \mathfrak{M}_{ij}$, so that $\dim \mathfrak{M} \geq \dim \mathfrak{M}_{ij}$. By Corollary 1 to Proposition 9, the set of elements of $\mathfrak{M}_{ij}$ which do not belong to $\mathfrak{K}\mathfrak{N}$ lies in an algebraic set properly contained in $\mathfrak{M}_{ij}$, that is, in a finite union of irreducible sets all of dimension $< \dim \mathfrak{M}_{ij}$. It follows that $\mathfrak{M} - \mathfrak{K}\mathfrak{N}$ is contained in a finite union of irreducible sets each of dimension $< \dim \mathfrak{M}$, so that (Proposition 15) $\mathfrak{K}\mathfrak{N} = \mathfrak{M}$, whence $\mathfrak{K}\mathfrak{N}$ is algebraic.

PROPOSITION 17. *Let $\mathfrak{K}$ be an algebraic subgroup of an algebraic group $\mathfrak{H}$ in $\mathfrak{G}$, and let $N(\mathfrak{K})$ be the normalisor of $\mathfrak{K}$ in $\mathfrak{H}$. Then $N(\mathfrak{K})$ is an algebraic group in $\mathfrak{G}$.*

Proof. The intersection $\mathfrak{N}$ of all the algebraic sets containing $N(\mathfrak{K})$ is an algebraic group (Proposition 14); we shall show that $N(\mathfrak{K}) = \mathfrak{N}$, thereby proving that $N(\mathfrak{K})$ is algebraic. Let $\mathfrak{N}_0$ be any component of $\mathfrak{N}$ and let σ_0 be a generic element of $\mathfrak{N}_0$. Suppose $\sigma_0\mathfrak{K}^*\sigma_0^{-1} \not\subseteq \mathfrak{K}^*$. Using Proposition 9 we see that there exists an element $\tau \varepsilon \mathfrak{K}$ such that $\sigma_0\tau\sigma_0^{-1} \notin \mathfrak{K}^*$. Letting $\eta_1, \cdots, \eta_n$ be elements of $\mathfrak{G}$ such that $\mathfrak{G} = \mathfrak{F}\langle\eta_1, \cdots, \eta_n\rangle$, we see that therefore there exists a $Q \varepsilon \mathfrak{G}\{w_1, \cdots, w_n\}$ which vanishes at $(\sigma\eta_1, \cdots, \sigma\eta_n)$ for every $\sigma \varepsilon \mathfrak{K}$ but which does not vanish at $(\sigma_0\tau\sigma_0^{-1}\eta_1, \cdots, \sigma_0\tau\sigma_0^{-1}\eta_n)$. By Corollary 1 to Proposition 9 there exists a $P \varepsilon \mathfrak{G}\{y_1, \cdots, y_n\}$ which does not vanish at $(\sigma_0\eta_1, \cdots, \sigma_0\eta_n)$ such that whenever σ'_0 is a specialization of σ_0 for which $P(\sigma'_0\eta_1, \cdots, \sigma'_0\eta_n) \neq 0$ then $Q(\sigma'_0\tau\sigma'^{-1}_0\eta_1, \cdots, \sigma'_0\tau\sigma'^{-1}_0\eta_n) \neq 0$, that is then $\sigma'_0\tau\sigma'^{-1}_0 \notin \mathfrak{K}$, so that $\sigma'_0 \notin N(\mathfrak{K})$. Thus $P(\sigma\eta_1, \cdots, \sigma\eta_n) = 0$ for every $\sigma \varepsilon N(\mathfrak{K}) \cap \mathfrak{N}_0$. It easily follows that $N(\mathfrak{K}) \cap \mathfrak{N}_0$ is contained in an algebraic

set properly contained in $\mathfrak{N}_0$, so that $N(\mathfrak{K})$ is contained in an algebraic set properly contained in $\mathfrak{N}$. This contradicts the definition of $\mathfrak{N}$, and proves that $\sigma_0\mathfrak{K}^*\sigma_0^{-1} \subseteq \mathfrak{K}^*$. From this it easily follows, using Corollary 1 to Proposition 9, that $\sigma\mathfrak{K}\sigma^{-1} \subseteq \mathfrak{K}$ for all $\sigma \varepsilon \mathfrak{N}$ save possibly those in a finite union of irreducible sets all of dimension $< \dim \mathfrak{N}$, that is $\mathfrak{N} - \mathfrak{N}(K)$ is contained in such a union. From Proposition 15 it now follows that $N(\mathfrak{K}) = \mathfrak{N}$.

Chapter III. Galois theory of strongly normal extensions.

We recall that the conditions and conventions set forth at the beginning of Chapter II remain in force in the present chapter.

1. Normal extensions. We recall (Kolchin [3], Chapter III) two definitions: 1) a set of isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ is said to be *abundant* if for every differential field $\mathcal{F}_1$ between $\mathcal{F}$ and $\mathcal{G}$ and every element α in $\mathcal{G}$ but not in $\mathcal{F}_1$ there exists an isomorphism, in the set, which leaves invariant each element of $\mathcal{F}_1$ but which does not leave α invariant; 2) $\mathcal{G}$ is said to be a *normal* extension of $\mathcal{F}$ if the set of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is abundant. We now introduce the following definition: $\mathcal{G}$ is said to be a *weakly normal* extension of $\mathcal{F}$ if for every element α in $\mathcal{G}$ but not in $\mathcal{F}$ there exists an automorphism of $\mathcal{G}$ over $\mathcal{F}$ which does not leave α invariant. It is clear that if $\mathcal{G}$ is normal over $\mathcal{F}$ then $\mathcal{G}$ is weakly normal over $\mathcal{F}$, and that $\mathcal{G}$ is normal over $\mathcal{F}$ if and only if $\mathcal{G}$ is weakly normal over every differential field between $\mathcal{F}$ and $\mathcal{G}$. Whether $\mathcal{G}$ can be weakly normal over $\mathcal{F}$ without being normal over $\mathcal{F}$ is an open question.[7]

The following result was proved in Kolchin [3], § 16.

THEOREM 1. *Let $\mathcal{G}$ be a normal extension of $\mathcal{F}$, let $\mathfrak{G}$ be an abundant group of automorphisms of $\mathcal{G}$ over $\mathcal{F}$* (not necessarily the group of all such automorphisms), *and for each differential field $\mathcal{F}_1$ between $\mathcal{F}$ and $\mathcal{G}$ let*

[7] If we relax the conditions on $\mathcal{G}$ and $\mathcal{F}$ by dropping the requirement that every constant in $\mathcal{G}$ belong to $\mathcal{F}$ then the answer to this question is affirmative, as is shown by the following example. Let $\mathcal{F}$ be the field of all algebraic numbers, let θ be a transcendental number, let $\mathcal{G} = \mathcal{F}(\theta)$, and make $\mathcal{F}$ and $\mathcal{G}$ into ordinary differential fields by defining $\delta_1 a = 0$ for every $a \varepsilon \mathcal{G}$. The automorphisms of $\mathcal{G}$ over $\mathcal{F}$ may be identified with the fractional linear substitutions $\theta \to (a\theta + b)(c\theta + d)^{-1}$ with $a, b, c\, d \varepsilon \mathcal{F}$ and $ad - bc \neq 0$. The only elements of $\mathcal{G}$ invariant under $\theta \to \theta + 1$ are those of $\mathcal{F}$, so that $\mathcal{G}$ is weakly normal over $\mathcal{F}$ (in the relaxed sense). But $\theta \not\varepsilon \mathcal{F}(\theta^3 + \theta)$, and the only fractional linear substitution leaving $\theta^3 + \theta$ invariant is $\theta \to \theta$, which leaves θ invariant, too; therefore $\mathcal{G}$ is not normal over $\mathcal{F}$ (in the relaxed sense).

$\mathfrak{G}(\mathcal{F}_1)$ denote the group of all elements of $\mathfrak{G}$ which leave invariant each element of $\mathcal{F}_1$. Then for each $\mathcal{F}_1$ the field of all elements of $\mathcal{G}$ invariant under every element of $\mathfrak{G}(\mathcal{F}_1)$ is $\mathcal{F}_1$, so that $\mathcal{F}_1 \to \mathfrak{G}(\mathcal{F}_1)$ is a one-to-one mapping of the set of all differential fields between $\mathcal{F}$ and $\mathcal{G}$ onto a certain set of subgroups of $\mathfrak{G}$. A necessary and sufficient condition that $\mathfrak{G}(\mathcal{F}_1)$ be a normal subgroup of $\mathcal{G}$ is that $\sigma\mathcal{F}_1 \subseteq \mathcal{F}_1$ for every $\sigma \varepsilon \mathfrak{G}$, and when this condition is satisfied then the mapping which to each element of $\mathfrak{G}$ assigns its restriction to $\mathcal{F}_1$ is a homomorphism with kernel $\mathfrak{G}(\mathcal{F}_1)$ of $\mathfrak{G}$ onto an abundant group of automorphisms of $\mathcal{F}_1$ over $\mathcal{F}$.

We shall give an example which will show that even when $\mathfrak{G}$ is taken as the group of *all* automorphisms of $\mathcal{G}$ over $\mathcal{F}$ and $\mathfrak{G}(\mathcal{F}_1)$ is a normal subgroup of $\mathfrak{G}$ (so that $\mathcal{F}_1$ is a normal extension of $\mathcal{F}$) the factor group $\mathfrak{G}/\mathfrak{G}(\mathcal{F}_1)$ need not be isomorphic to the group of all automorphisms of $\mathcal{F}_1$ over $\mathcal{F}$, but merely to an abundant subgroup thereof; the example will also show that it is possible for an intermediate differential field $\mathcal{F}_1$ that $\mathcal{F}_1$ be a normal extension of $\mathcal{F}$ and $\mathfrak{G}(\mathcal{F}_1)$ fail to be a normal subgroup of $\mathfrak{G}$. To discuss this example we shall need the following lemma.

LEMMA 1. *Let p, q be integers not both zero, let a, b, c, d be nonzero elements of $\mathcal{C}$, let X_1, X_2, X_3, X_4 be indeterminates, and let $f \varepsilon \mathcal{C}(X_1, X_2, X_3, X_4)$. If f is invariant under the substitution*

$$(X_1, X_2, X_3, X_4) \to (X_1 + d, aX_2, bX_3, cX_2{}^pX_3{}^qX_4) \tag{1}$$

then $f \varepsilon \mathcal{C}(X_2, X_3)$.

Proof. We suppose as we may that $f \neq 0$; then f is uniquely expressible in the form

$$f = \sum_{i=0}^{m} g_iX_4{}^i / \sum_{j=0}^{m} h_jX_4{}^j \, (g_i, h_j \varepsilon \mathcal{C}(X_1, X_2, X_3), g_m \neq 0, h_n = 1),$$

where $\sum g_iX_4{}^i$ and $\sum h_jX_4{}^j$ have no common factor as polynomials in X_4. Because of the invariance of f under the indicated substitution we find that

$$g_i(X_1 + d, aX_2, bX_3) = g_i(X_1, X_2, X_3)\,(cX_2{}^pX_3{}^q)^{n-i},$$

$$h_j(X_1 + d, aX_2, bX_3) = h_j(X_1, X_2, X_3)\,(cX_2{}^pX_3{}^q)^{n-j}.$$

But the degree in X_2 of the numerator of any nonzero element $\phi \varepsilon \mathcal{C}(X_1, X_2, X_3)$ minus the degree in X_2 of the denominator of ϕ is obviously invariant under the substitution

$$(X_1, X_2, X_3) \to (X_1 + d, aX_2, bX_3); \tag{2}$$

therefore $p(n-i)=0$ for every i such that $g_i \neq 0$. Similarly, regarding degrees in X_3 instead of X_2, we see that $q(n-i)=0$ whenever $g_i \neq 0$. Since p and q are not both 0 this implies that $g_i = 0$ whenever $i \neq n$. In the same way we find that $h_j = 0$ whenever $j \neq n$. Therefore $m = n = 0$, and $f \varepsilon \mathcal{C}(X_1, X_2, X_3)$; thus we may write $f = f(X_1, X_2, X_3)$, and f is invariant under the substitution (2).

Now the set $\mathfrak{H}$ of all matrices

$$\tau(\alpha,\beta,\delta) = \begin{bmatrix} 1 & \delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & \beta \end{bmatrix} \quad (\alpha, \beta, \delta \,\varepsilon\, \mathcal{C}, \alpha\beta \neq 0)$$

such that $f(X_1+\delta, \alpha X_2, \beta X_3) = f(X_1, X_2, X_3)$, is an algebraic matic group, and $\tau(a, b, d) \,\varepsilon\, \mathfrak{H}$. It follows (Kolchin [5], § 3, Lemma 1) that $\tau(1, 1, d) \,\varepsilon\, \mathfrak{H}$, that is, $f(X_1+d, X_2, X_3) = f(X_1, X_2, X_3)$. Since $d \neq 0$ it is easy to conclude that f is free of X_1, that is $f \,\varepsilon\, \mathcal{C}(X_2, X_3)$.

EXAMPLE. Let $\mathcal{F}$ be the field of all complex numbers and let $\mathcal{G}$ be the field $\mathcal{F}(x, e^x, e^{ix}, e^{x^2})$ obtained by the field adjunction to $\mathcal{F}$ of the functions x, e^x, e^{ix}, e^{x^2} of the complex variable x $(i = \sqrt{-1})$; with respect to the operator d/dx, $\mathcal{G}$ is an ordinary differential field with field of constants $\mathcal{F}$.

We shall first show that if $\mathcal{F}_1$ is a differential field between $\mathcal{F}$ and $\mathcal{G}$ then either $\mathcal{F}_1$ is between $\mathcal{F}\langle x\rangle$ and $\mathcal{G}$ or else $\mathcal{F}_1$ is between $\mathcal{F}$ and $\mathcal{F}\langle e^x, e^{ix}\rangle$. Indeed, suppose $x \not\varepsilon\, \mathcal{F}_1$, and let $\theta \,\varepsilon\, \mathcal{F}_1$; we must prove that $\theta \,\varepsilon\, \mathcal{F}\langle e^x, e^{ix}\rangle$. Since $x \not\varepsilon\, \mathcal{F}_1$ there exists an isomorphism of $\mathcal{G}$ over $\mathcal{F}_1$ under which x is not invariant; it easily follows that there exist nonzero complex numbers a_0, b_0, c_0, d such that $(x+d, a_0e^x, b_0e^{ix}, c_0e^{(x+d)^2})$ is a generic specialization of $(x, e^x, e^{ix}, e^{x^2})$ over $\mathcal{F}_1$. Letting $c = c_0e^{d^2}$ and letting f be an element of $\mathcal{F}(X_1, X_2, X_3, X_4)$ such that $\theta = f(x, e^x, e^{ix}, e^{x^2})$, we see that

$$f(x+d, a_0e^x, b_0e^{ix}, ce^{2dx}e^{x^2}) = f(x, e^x, {}^{ix}, e^{x^2}). \tag{3}$$

If f were not free of X_4 this would mean that $e^x, e^{ix}, e^{2dx}, e^{x^2}$ are algebraically dependent over $\mathcal{F}(x)$, and this would imply that $x, ix, 2dx, x^2$ are linearly dependent over the ring of integers, that is, there would exist integers p, q, r with $r \neq 0$ and p, q not both 0 such that $2d = (p+iq)/r$. Choosing complex numbers a, b such that $a_0 = a^r$, $b_0 = b^r$ we would then have

$$f(x+d, a^re^x, b^re^{ix}, ce^{[(p+iq)/r]x}e^{x^2}) = f(x, e^x, e^{ix}, e^{x^2}).$$

Since $e^{x/r}, e^{ix/r}, e^{x^2}$ are algebraically independent over $\mathcal{F}(x)$, this would imply

that $f(X_1, X_2{}^r, X_3{}^r, X_4)$ is invariant under the substitution (1), so that by Lemma 1 we would have $f \varepsilon \mathfrak{F}(X_2, X_3)$. It follows that f is free of X_4. Therefore we may write (3) in the form

$$f(x+d, a_0 e^x, b_0{}^{ix}, c_0 e^x e^{x^2}) = f(x, e^x, e^{ix}, e^{x^2}),$$

whence f is invariant under the substitution (1) (with $(a_0, b_0, c_0, d, 1, 0)$ instead of (a, b, c, d, p, q)). It follows from Lemma 1 that $f \varepsilon \mathfrak{F}(X_2, X_3)$, so that $\theta \varepsilon \mathfrak{F}\langle e^x, e^{ix}\rangle$. This completes the proof that $\mathfrak{F}_1$ is either between $\mathfrak{F}\langle x\rangle$ and $\mathfrak{G}$ or between $\mathfrak{F}$ and $\mathfrak{F}\langle e^x, e^{ix}\rangle$.

It is easy to see that for each choice of nonzero complex numbers a, b, c and integers p, q there exists a unique automorphism $\sigma = \sigma(p, q, a, b, c)$ of $\mathfrak{G}$ over $\mathfrak{F}$ such that $\sigma x = x + \frac{1}{2}(p + iq)$, $\sigma e^x = ae^x$, $\sigma e^{ix} = be^{ix}$, $\sigma e^{x^2} = ce^{px}e^{qix}e^{x^2}$, and every automorphism of $\mathfrak{G}$ over $\mathfrak{F}$ is of this form; let us denote the group of all these automorphisms by $\mathfrak{S}$.

Let $\mathfrak{F}_1$ be between $\mathfrak{F}$ and $\mathfrak{G}$. If $\mathfrak{F}_1$ is between $\mathfrak{F}\langle x\rangle$ and $\mathfrak{G}$ then, since $\mathfrak{G} = \mathfrak{F}\langle x\rangle\langle e^x, e^{ix}, e^{x^2}\rangle$ is obviously a Picard-Vessiot extension of $\mathfrak{F}\langle x\rangle$, $\mathfrak{G}$ is normal over $\mathfrak{F}\langle x\rangle$ and therefore weakly normal over $\mathfrak{F}_1$. Suppose, then, that $\mathfrak{F}_1$ is between $\mathfrak{F}$ and $\mathfrak{F}\langle e^x, e^{ix}\rangle$, and let α be an element of $\mathfrak{G}$ but not of $\mathfrak{F}_1$. If $\alpha \not\varepsilon \mathfrak{F}_1\langle e^x, e^{ix}\rangle$ then, by Lemma 1 and the algebraic independence of e^x, e^{ix}, e^{x^2} over $\mathfrak{F}(x)$, the automorphism $\sigma(1, 0, 1, 1, 1)$ does not leave α invariant but obviously leaves each element of $\mathfrak{F}_1$ invariant; suppose, then, that $\alpha \varepsilon \mathfrak{F}\langle e^x, e^{ix}\rangle$. Since $\mathfrak{F}\langle e^x, e^{ix}\rangle$ is obviously a Picard-Vessiot extension of $\mathfrak{F}$, there exists an automorphism σ_0 of $\mathfrak{F}\langle e^x, e^{ix}\rangle$ over $\mathfrak{F}_1$ such that $\sigma_0\alpha \neq \alpha$; now there exist nonzero complex numbers a, b such that $\sigma_0 e^x = ae^x$, $\sigma_0 e^{ix} = be^{ix}$. It is clear that the element $\sigma(0, 0, a, b, 1)$ of $\mathfrak{S}$ is an extension of σ_0, and therefore does not leave α invariant but leaves each element of $\mathfrak{F}_1$ invariant. Thus again $\mathfrak{G}$ is a weakly normal extension of $\mathfrak{F}_1$. Since $\mathfrak{F}_1$ is arbitrary between $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{G}$ is a normal extension of $\mathfrak{F}$.

Now let $\mathfrak{F}_2 = \mathfrak{F}\langle x, e^x, e^{x^2}\rangle$, $\mathfrak{F}_3 = \mathfrak{F}\langle x\rangle$. It is clear that $\sigma\mathfrak{F}_3 = \mathfrak{F}_3$ for every $\sigma = \sigma(p, q, a, b, c) \varepsilon \mathfrak{S}$, so that $\mathfrak{S}(\mathfrak{F}_3)$ is a normal subgroup of $\mathfrak{S}$; the restrictions to $\mathfrak{F}_3$ of the elements of $\mathfrak{S}$ are the automorphisms $\sigma(d)$ of $\mathfrak{F}_3$ over $\mathfrak{F}$ defined by

$$\sigma(d)x = x + d,$$

with d an arbitrary number of the form $\frac{1}{2}(p + iq)$, p and q being integers, and the group of all these restrictions is not the group of all automorphisms of $\mathfrak{F}_3$ over $\mathfrak{F}$, that is, is not the group of all automorphisms $\sigma(d)$ with d an arbitrary complex number, but is merely an abundant subgroup thereof.

Finally, there exists an automorphism $\sigma \varepsilon \mathfrak{S}$ such that $\sigma F_2 \not\subset F_2$ (for

example $\sigma = \sigma(0, 1, 1, 1, 1)$), so that $\mathfrak{G}(\mathcal{F}_2)$ is not a normal subgroup of $\mathfrak{G}$; nevertheless $\mathcal{F}_2$ is a normal extension of $\mathcal{F}$, as can be shown by the method used to prove that $\mathcal{G}$ is a normal extension of $\mathcal{F}$.

2. Strongly normal extensions. We make the following definition: $\mathcal{G}$ is said to be a *strongly normal* extension of $\mathcal{F}$ if every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is strong. It is obvious that if $\mathcal{G}$ is strongly normal over $\mathcal{F}$ then $\mathcal{G}$ is strongly normal over every differential field between $\mathcal{F}$ and $\mathcal{G}$.

PROPOSITION 1. *A necessary and sufficient condition that $\mathcal{G}$ be a strongly normal extension of $\mathcal{F}$ is that $\chi\mathcal{G} \subseteq \mathcal{G}\langle\mathcal{E}^*\rangle$ for every isolated isomorphism χ of $\mathcal{G}$ over $\mathcal{F}$.*

Proof. That the condition is necessary is obvious; suppose then that the condition is satisfied. The differential field $\chi\mathcal{G}$, isomorphic with $\mathcal{G}$, also must have this property, so that $\mathcal{G} = \chi^{-1}(\chi\mathcal{G}) \subset (\chi\mathcal{G})\langle\mathcal{E}^*\rangle$, and χ is strong. It follows (Chapter II, Propositions 2 and 5) that every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is strong, so that $\mathcal{G}$ is strongly normal over $\mathcal{F}$.

PROPOSITION 2. *If $\mathcal{G}$ is strongly normal over $\mathcal{F}$ then $\mathcal{G}$ is normal over $\mathcal{F}$.*

Proof. Let $\mathcal{F}_1$ be a differential field between $\mathcal{F}$ and $\mathcal{G}$, and let α be an element of $\mathcal{G}$ not in $\mathcal{F}_1$. There exists an isomorphism σ of $\mathcal{G}$ over $\mathcal{F}_1$ such that $\sigma\alpha - \alpha \neq 0$, and by hypothesis σ is strong. By Corollary 2 to Proposition 9 of Chapter II there exists an automorphism τ of $\mathcal{G}$ over $\mathcal{F}_1$ which is a specialization of σ such that $\tau\alpha - \alpha \neq 0$. Therefore $\mathcal{G}$ is normal over $\mathcal{F}$.

That $\mathcal{G}$ can be normal over $\mathcal{F}$ without being strongly normal over $\mathcal{F}$ is shown by the example of § 1; using the notation of that example we easily see that for each complex number d there exists a unique isomorphism σ of $\mathcal{G}$ over $\mathcal{F}$ such that $\sigma x = x + d$, $\sigma e^x = e^d e^x$, $\sigma e^{ix} = e^{id} e^{ix}$, $\sigma e^{x^2} = e^{d^2} e^{2dx} e^{x^2}$, and it is obvious that if $2d$ is not a gaussian integer then $e^{2dx} \notin \mathcal{G} = \mathcal{F}\langle x, e^x, e^{ix}, e^{x^2}\rangle$ so that $\mathcal{G}\langle\sigma\mathcal{G}\rangle \not\subseteq \mathcal{G}\langle\mathcal{E}_\sigma\rangle = \mathcal{G}$, whence $\mathcal{G}$ is not strongly normal over $\mathcal{F}$.

PROPOSITION 3. *If $\mathcal{G}$ is a strongly normal extension of $\mathcal{F}$ then the group of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is algebraic.*

Proof. By Proposition 1 of Chapter II there exists a finite number of isomorphisms $\chi_1, \cdots, \chi_h$ of $\mathcal{G}$ over $\mathcal{F}$ such that every isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is a specialization of one of these; by hypothesis χ_i is strong and is

therefore a generic element of an irreducible set $\mathfrak{M}_i$ of automorphisms of $\mathcal{G}$ over $\mathcal{F}$. The group of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is $\bigcup \mathfrak{M}_i$ and therefore is algebraic.

3. The fundamental theorems. Whenever $\mathcal{G}$ is strongly normal over $\mathcal{F}$ we shall denote the group of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ by $\mathfrak{G}$, and for each intermediate differential field $\mathcal{F}_1$ we shall denote the group of all automorphisms of $\mathcal{G}$ over $\mathcal{F}_1$ by $\mathfrak{G}(\mathcal{F}_1)$.

THEOREM 2. *If $\mathcal{G}$ is strongly normal over $\mathcal{F}$, then the mapping $\mathcal{F}_1 \to \mathfrak{G}(\mathcal{F}_1)$ is one-to-one from the set of all differential fields between $\mathcal{F}$ and $\mathcal{G}$ onto the set of all algebraic groups in $\mathfrak{G}$, and has the property that* $\dim \mathfrak{G}(\mathcal{F}_1) = \partial^0 \mathcal{G}/\mathcal{F}_1$.

Proof. By Proposition 2 and Theorem 1 the mapping is one-to-one, and by Proposition 3 $\mathfrak{G}(\mathcal{F}_1)$ is algebraic; a generic element of a component of $\mathfrak{G}(\mathcal{F}_1)$ is an isolated isomorphism of $\mathcal{G}$ over $\mathcal{F}_1$ whence (Chapter II, Proposition 2) $\dim \mathfrak{G}(\mathcal{F}_1) = \partial^0 \mathcal{G}/\mathcal{F}_1$. It remains to prove the mapping is onto. To this end let $\mathfrak{G}_1$ be any algebraic group in $\mathfrak{G}$, and let $\mathcal{F}_1$ be the differential field of invariants of $\mathfrak{G}_1$; we shall show that $\mathfrak{G}_1 = \mathfrak{G}(\mathcal{F}_1)$, thereby proving that $\mathfrak{G}_1$ is algebraic and completing the proof of the theorem. Now, it is obvious that $\mathfrak{G}_1 \subseteq \mathfrak{G}(\mathcal{F}_1)$. Suppose that $\mathfrak{G}_1 \neq \mathfrak{G}(\mathcal{F}_1)$. Then, if $\eta_1, \cdots, \eta_n$ are elements such that $\mathcal{G} = \mathcal{F}\langle \eta_1, \cdots, \eta_n \rangle$, there exists a differential polynomial in $\mathcal{G}\{y_1, \cdots, y_n\}$ which vanishes at

$$(\sigma\eta_1, \cdots, \sigma\eta_n) \tag{4}$$

for every $\sigma \in \mathfrak{G}_1$ but not for every $\sigma \in \mathfrak{G}(\mathcal{F}_1)$; of all such differential polynomials let F be one with a minimal number of terms, and assume without loss of generality that one of the coefficients in F is 1. Let τ be any element of $\mathfrak{G}_1$ and let F_τ be the differential polynomial obtained when each coefficient ϕ in F is replaced by $\tau\phi$. Then

$$F_\tau(\sigma\eta_1, \cdots, \sigma\eta_n) = \tau F(\tau^{-1}\sigma\eta_1, \cdots, \tau^{-1}\sigma\eta_n) = 0$$

for every $\sigma \in \mathfrak{G}_1$, so that $F - F_\tau$ vanishes at (4) for every $\sigma \in \mathfrak{G}_1$. Now $F - F_\tau$ has fewer terms than F has, so that $F - F_\tau$ must vanish at (4) for every $\sigma \in \mathfrak{G}(\mathcal{F}_1)$. If $F - F_\tau$ were not 0 there would exist an element $\gamma \in \mathcal{G}$ such that $F - \gamma(F - F_\tau)$ has fewer terms than F has; but $F - \gamma(F - F_\tau)$ obviously vanishes at (4) for every $\sigma \in \mathfrak{G}_1$ but not for every $\sigma \in \mathfrak{G}(\mathcal{F}_1)$, and therefore can not have fewer terms that F has. It follows that $F - F_\tau = 0$. Since this is true for every $\tau \in \mathfrak{G}_1$, each coefficient in F belongs to F_1. From

this it follows that F vanishes at (4) for every $\sigma \varepsilon \mathfrak{G}(\mathcal{F}_1)$. This contradiction proves that $\mathfrak{G}_1 = \mathfrak{G}(\mathcal{F}_1)$, and completes the proof of the theorem.

THEOREM 3. *If $\mathcal{G}$ is strongly normal over $\mathcal{F}$, $\mathcal{F}_1$ is a differential field between $\mathcal{F}$ and $\mathcal{G}$, and σ_1 is an isomorphism of $\mathcal{F}_1$ over $\mathcal{F}$ into $\mathcal{G}$, then there exists an automorphism $\sigma \varepsilon \mathfrak{G}$ which is an extension of σ_1.*

Proof. σ_1 can be extended to an isomorphism σ' of $\mathcal{G}$ over $\mathcal{F}$. By Corollary 2 to Proposition 9 of Chapter II there exist an automorphism σ of $\mathcal{G}$ which is a specialization of σ' and which therefore coincides with σ_1 on $\mathcal{F}_1$.

THEOREM 4. *Let $\mathcal{G}$ be a strongly normal extension of $\mathcal{F}$ and let $\mathcal{F}_1$ be a differential field between $\mathcal{F}$ and $\mathcal{G}$. Then the following five conditions are equivalent.* 1) *$\mathcal{F}_1$ is strongly normal over $\mathcal{F}$*; 2) *$\mathcal{F}_1$ is normal over $\mathcal{F}$*; 3) *$\mathcal{F}_1$ is weakly normal over $\mathcal{F}$*; 4) *$\sigma\mathcal{F}_1 \subseteq \mathcal{F}_1$ for every $\sigma \varepsilon \mathfrak{G}$*; 5) *$\mathfrak{G}(\mathcal{F}_1)$ is a normal subgroup of $\mathfrak{G}$. When these conditions are satisfied then the mapping, which to each $\sigma \varepsilon \mathfrak{G}$ assigns the restriction of σ to $\mathcal{F}_1$, is a homomorphism with kernel $\mathfrak{G}(\mathcal{F}_1)$ of $\mathfrak{G}$ onto the group of all automorphisms of $\mathcal{F}_1$ over $\mathcal{F}$.*

Proof. We already know that 1) implies 2) and that 2) implies 3); also, by Theorem 1, 4) is equivalent to 5). To prove the equivalence of the five conditions it suffices to show that 3) implies 5) and that 4) implies 1). To settle the first point suppose that $\mathfrak{G}(\mathcal{F}_1)$ is not a normal subgroup of $\mathfrak{G}$, so that the normalisor $N(\mathfrak{G}(\mathcal{F}_1)) \neq \mathfrak{G}$; by Proposition 17 of Chapter II, $N(\mathfrak{G}(\mathcal{F}_1))$ is an algebraic group in $\mathfrak{G}$, so that by Theorem 2 there exists a differential field $\mathcal{F}_2$ between $\mathcal{F}$ and $\mathcal{F}_1$ with $\mathcal{F}_2 \neq \mathcal{F}$, such that $N(\mathfrak{G}(\mathcal{F}_1)) = \mathfrak{G}(\mathcal{F}_2)$. Let α be any element of $\mathcal{F}_2$ not in $\mathcal{F}$. If σ_1 is any automorphism of $\mathcal{F}_1$ over $\mathcal{F}$ then (Theorem 3) σ_1 can be extended to an automorphism $\sigma \varepsilon \mathfrak{G}$; since $\sigma\mathcal{F}_1 = \sigma_1\mathcal{F}_1 = \mathcal{F}_1$, we see that $\tau\sigma\beta = \sigma\beta$ for every $\beta \varepsilon \mathcal{F}_1$ and every $\tau \varepsilon \mathfrak{G}(\mathcal{F}_1)$, that is $\sigma^{-1}\tau\sigma\beta = \beta$, so that $\sigma^{-1}\tau\sigma \varepsilon \mathfrak{G}(\mathcal{F}_1)$, whence $\sigma \varepsilon N(\mathfrak{G}(\mathcal{F}_1)) = \mathfrak{G}(\mathcal{F}_2)$. Therefore $\sigma_1\alpha = \sigma\alpha = \alpha$, that is, every automorphism of $\mathcal{F}_1$ over $\mathcal{F}$ leaves α invariant, so that $\mathcal{F}_1$ is not weakly normal over $\mathcal{F}$. Thus, if $\mathcal{F}_1$ is weakly normal over $\mathcal{F}$ then $\mathfrak{G}(\mathcal{F}_1)$ is a normal subgroup of $\mathfrak{G}$.

To settle the second point, suppose that $\mathcal{F}_1$ is not strongly normal over $\mathcal{F}$. Then (Proposition 1) there exists an isomorphism σ_1 of $\mathcal{F}_1$ over $\mathcal{F}$ such that $\sigma_1\mathcal{F}_1 \not\subseteq \mathcal{F}_1\langle \mathcal{C}^* \rangle$, and σ_1 can be extended to an isomorphism σ of $\mathcal{G}$ over $\mathcal{F}$. Let θ be an element of $\sigma_1\mathcal{F}_1 = \sigma\mathcal{F}_1$ which does not belong to

$\mathcal{F}_1\langle\mathcal{C}^*\rangle$. We claim there exists a $\tau \in \mathfrak{G}(\mathcal{F}_1)$ such that $\tau\theta \neq \theta$. Indeed, since $\theta \in \sigma\mathcal{F}_1 \subseteq \mathcal{G}\langle\sigma\mathcal{G}\rangle = \mathcal{G}\langle\mathcal{C}_\sigma\rangle$, we may write

$$\theta B(c_1, \cdots, c_r) = \sum_{i=0}^{s-1} A_i(c_1, \cdots, c_r) d^i,$$

where $B, A_0, \cdots, A_{s-1}$ are polynomials in $\mathcal{G}[u_1, \cdots, u_r]$ without common divisor, one of the coefficients in B is 1, $c_1, \cdots, c_r$ are elements of $\mathcal{C}_\sigma$ algebraically independent over $\mathcal{C}$ (and hence over $\mathcal{G}$), and d is an element of $\mathcal{C}_\sigma$ which is algebraic of some degree s over $\mathcal{G}\langle c_1, \cdots, c_r\rangle$. If $\tau \in \mathfrak{G}(\mathcal{F}_1)$ has the property that $\tau\theta = \theta$ then,

$$\sum_{i=0}^{s-1} (B_\tau(c_1, \cdots, c_r) A_i(c_1, \cdots, c_r) - B(c_1, \cdots, c_r) A_{i\tau}(c_1, \cdots, c_r)) d^i = 0$$

(where in general for any polynomial C we denote by C_τ the polynomial obtained upon replacing each coefficient in C by its image under τ), whence $B_\tau A_i = BA_{i\tau}$ $(0 \leq i \leq s-1)$. Because $B, A_0, \cdots, A_{s-1}$ have no common divisor and one of the coefficients in B is 1, it follows that $B_\tau = B$, $A_{i\tau} = A_i$ $(0 \leq i \leq s-1)$. Therefore if θ were invariant under every $\tau \in \mathfrak{G}(\mathcal{F}_1)$ then so would every coefficient in B and each A_i, and these coefficients would all belong to $\mathcal{F}_1$, contradicting the fact that $\theta \notin \mathcal{F}_1\langle\mathcal{C}^*\rangle$. This establishes our claim that for some $\tau \in \mathfrak{G}(\mathcal{F}_1)$ we have $\tau\theta \neq \theta$.

Now $\theta \in \sigma\mathcal{F}_1$, so that there exists an element $\zeta \in \mathcal{F}_1$ such that $\theta = \sigma\zeta$. For this ζ and the above τ we have $\tau\sigma\zeta \neq \sigma\zeta$. By Chapter II, Proposition 9, there exists an automorphism σ_0 of $\mathcal{G}$ which is a specialization of σ for which $\tau\sigma_0\zeta \neq \sigma_0\zeta$. Since $\tau \in \mathfrak{G}(\mathcal{F}_1)$ this means that $\sigma_0\zeta \notin \mathcal{F}_1$ and since $\zeta \in \mathcal{F}_1$ this shows that $\sigma_0\mathcal{F}_1 \not\subseteq \mathcal{F}_1$. We have thus shown that if $\mathcal{F}_1$ is not strongly normal over $\mathcal{F}$ then $\sigma_0\mathcal{F}_1 \not\subseteq \mathcal{F}_1$ for some $\sigma_0 \in \mathfrak{G}$, so that 4) implies 1). This completes the proof that the five conditions are equivalent.

When these conditions are satisfied then (Theorem 1) the mapping which to each automorphism in $\mathfrak{G}$ assigns its restriction to $\mathcal{F}_1$ is a homomorphism with kernel $\mathfrak{G}(\mathcal{F}_1)$ of $\mathfrak{G}$ *into* the group of all automorphisms of $\mathcal{F}_1$ over $\mathcal{F}$. That this homomorphism is *onto* follows from Theorem 3.

COROLLARY. *If $\mathcal{G}$ and $\mathcal{H}$ are strongly normal extension of $\mathcal{F}$ such that the field of constants of $\mathcal{F}\langle\mathcal{G}, \mathcal{H}\rangle$ is $\mathcal{C}$, then $\mathcal{F}\langle\mathcal{G}, \mathcal{H}\rangle$ and $\mathcal{G} \cap \mathcal{H}$ are strongly normal over $\mathcal{F}$.*

Proof. Let σ be any isomorphism of $\mathcal{F}\langle\mathcal{G}, \mathcal{H}\rangle$ over $\mathcal{F}$; the restrictions of σ to $\mathcal{G}$ and to $\mathcal{H}$ are isomorphisms of $\mathcal{G}$ and of $\mathcal{H}$ over $\mathcal{F}$ and, since $\mathcal{G}$ and $\mathcal{H}$ are strongly normal over $\mathcal{F}$, we have

$$\sigma(\mathcal{F}\langle\mathcal{G}, \mathcal{H}\rangle) = \mathcal{F}\langle\sigma\mathcal{G}, \sigma\mathcal{H}\rangle$$
$$\subseteq \mathcal{F}\langle\mathcal{G}\langle\mathcal{C}^*\rangle, \mathcal{H}\langle\mathcal{C}^*\rangle\rangle = \mathcal{F}\langle\mathcal{G}, \mathcal{H}\rangle\langle\mathcal{C}^*\rangle.$$

Therefore (Proposition 1) $\mathcal{F}\langle\mathcal{G},\mathcal{H}\rangle$ is strongly normal over $\mathcal{F}$. Now let τ be any automorphism of $\mathcal{F}\langle\mathcal{G},\mathcal{H}\rangle$ over $\mathcal{F}$. Since $\mathcal{G}$ and $\mathcal{H}$ are both strongly normal over $\mathcal{F}$ we see by Theorem 4 that $\tau\mathcal{G}\subseteq\mathcal{G}$ and $\tau\mathcal{H}\subseteq\mathcal{H}$, whence $\tau(\mathcal{G}\cap\mathcal{H})\subseteq\mathcal{G}\cap\mathcal{H}$, so that (Theorem 4) $\mathcal{G}\cap\mathcal{H}$ is strongly normal over $\mathcal{F}$.

THEOREM 5. *Let $\mathfrak{m}$ be a set of elements such that the field of constants of $\mathcal{G}\langle\mathfrak{m}\rangle$ is $\mathcal{C}$, let $\mathcal{F}^\dagger=\mathcal{F}\langle\mathfrak{m}\rangle$ and $\mathcal{G}^\dagger=\mathcal{G}\langle\mathfrak{m}\rangle$, and denote the group of all automorphisms of $\mathcal{G}^\dagger$ over $\mathcal{F}^\dagger$ by $\mathfrak{G}^\dagger$. If $\mathcal{G}$ is strongly normal over $\mathcal{F}$ then $\mathfrak{G}^\dagger$ is strongly normal over $\mathcal{F}^\dagger$, and the mapping which to each $\sigma^\dagger\in\mathfrak{G}^\dagger$ assigns the restriction σ of $\sigma^\dagger$ to $\mathcal{G}$ is an isomorphism of $\mathfrak{G}^\dagger$ onto $\mathfrak{G}(\mathcal{F}^\dagger\cap\mathcal{G})$ which maps every algebraic group $\mathfrak{M}^\dagger$ in $\mathfrak{G}^\dagger$ onto an algebraic group $\mathfrak{M}$ in $\mathfrak{G}$ of the same dimension; if $\mathfrak{M}^\dagger$ is irreducible then so is $\mathfrak{M}$.*

Proof. Let $\rho^\dagger$ be any isomorphism of $\mathcal{G}^\dagger$ over $\mathcal{F}^\dagger$ and let ρ be the restriction of $\rho^\dagger$ to $\mathcal{G}$. Then ρ is an isomorphism of $\mathcal{G}$ over $\mathcal{F}$, and

$$\mathcal{G}^\dagger\langle\rho^\dagger\mathcal{G}^\dagger\rangle=\mathcal{G}\langle\mathfrak{m}\rangle\langle(\rho\mathcal{G})\langle\mathfrak{m}\rangle\rangle=\mathcal{G}\langle\mathcal{C}_\rho\rangle\langle\mathfrak{m}\rangle=\mathcal{G}^\dagger\langle\mathcal{C}_\rho\rangle;$$

it follows (Proposition 1) that $\mathcal{G}^\dagger$ is strongly normal over $\mathcal{F}^\dagger$, and also (Corollary 5 to Proposition 3 of Chapter 1) that the field of constants of $\mathcal{G}^\dagger\langle\rho^\dagger\mathcal{G}^\dagger\rangle$ is $\mathcal{C}_\rho$.

The mapping $\sigma^\dagger\to\sigma$ is obviously a homomorphism of $\mathfrak{G}^\dagger$ into $\mathfrak{G}$. If σ is the identity automorphism of $\mathcal{G}$ then $\sigma^\dagger$ is obviously the identity automorphism of $\mathfrak{G}^\dagger$; therefore this homomorphism is an isomorphism of $\mathfrak{G}^\dagger$ onto some subgroup $\mathfrak{G}_1$ of $\mathfrak{G}$. An element $\alpha\in\mathcal{G}$ is invariant under every $\sigma\in\mathfrak{G}_1$, that is under every $\sigma^\dagger\in\mathfrak{G}^\dagger$, if and only if $\alpha\in\mathcal{F}^\dagger$; it follows that $\mathfrak{G}(\mathcal{F}^\dagger\cap\mathcal{G})$ is the smallest algebraic group in $\mathfrak{G}$ containing $\mathfrak{G}_1$, and therefore (Chapter II, Proposition 14) also the smallest algebraic set containg $\mathfrak{G}_1$.

Now let $\mathfrak{M}^\dagger$ be any irreducible set in $\mathfrak{G}^\dagger$, let $\rho^\dagger$ be a generic element of $\mathfrak{M}^\dagger$, and let ρ be the restriction of $\rho^\dagger$ to $\mathcal{G}$. Let $\mathfrak{M}$ be the set of all restrictions to $\mathcal{G}$ of elements of $\mathfrak{M}^\dagger$, and let $\mathfrak{M}_0$ be the irreducible set in $\mathfrak{G}$ with generic element ρ. Because every element of $\mathfrak{M}^\dagger$ is a specialization of $\rho^\dagger$, every element of $\mathfrak{M}$ is a specialization of ρ, so that $\mathfrak{M}\subseteq\mathfrak{M}_0$. If $\gamma_1,\cdots,\gamma_q$ are constants such that $\mathcal{C}_\rho=\mathcal{C}\langle\gamma_1,\cdots,\gamma_q\rangle$ then, by Proposition 9 of Chapter II and the fact that the field of constants of $\mathcal{G}^\dagger\langle\rho^\dagger\mathcal{G}^\dagger\rangle$ is $\mathcal{C}_\rho$, we know that there exists a polynomial $M\in\mathcal{C}[u_1,\cdots,u_q]$ with $M(\gamma_1,\cdots,\gamma_q)\neq 0$ such that whenever $c_1,\cdots,c_q$ are constants with $M(c_1,\cdots,c_q)\neq 0$ then there exists a unique isomorphism $\tau^\dagger$ of $\mathcal{G}^\dagger$ over $\mathcal{F}^\dagger$ for which $((\tau^\dagger\alpha)_{\alpha\in G^\dagger},c_1,\cdots,c_q)$ is a specialization of $((\rho^\dagger\alpha)_{\alpha\in G^\dagger},\gamma_1,\cdots,\gamma_q)$ over $\mathcal{G}^\dagger$. But it is easy to see that if $\eta_1,\cdots,\eta_n$ are elements of $\mathcal{G}$ such that $\mathcal{G}=\mathcal{F}\langle\eta_1,\cdots,\eta_n\rangle$, then there exists a differential polynomial

$K \varepsilon \mathscr{G}\{y_1, \cdots, y_n\}$ with $K(\rho\eta_1, \cdots, \rho\eta_n) \neq 0$ which has the following property: whenever τ is an isomorphism of $\mathscr{G}$ over $\mathscr{F}$ which is a specialization of ρ with $K(\tau\eta_1, \cdots, \tau\eta_n) \neq 0$ then there exist unique constants $c_1, \cdots, c_q$ with $M(c_1, \cdots, c_q) \neq 0$ such that $((\tau\alpha)_{\alpha \varepsilon G}, c_1, \cdots, c_q)$ is a specialization of $((\rho\alpha)_{\alpha \varepsilon G}, \gamma_1, \cdots, \gamma_q)$ over $\mathscr{G}$. It follows that every specialization τ of ρ such that $K(\tau\eta_1, \cdots, \tau\eta_n) \neq 0$ is the restriction to $\mathscr{G}$ of a specialization $\tau^\dagger$ of $\rho^\dagger$, and that $\tau^\dagger$ is an automorphism of $\mathscr{G}^\dagger$ if τ is an automorphism of $\mathscr{G}$. From this it is not difficult to conclude that $\mathfrak{M}_0 - \mathfrak{M}$ is contained in a finite union of irreducible sets in $\mathfrak{G}$ of lower dimension that $\mathfrak{M}_0$.

Suppose now that, in addition to being an irreducible set in $\mathfrak{G}^\dagger$, $\mathfrak{M}^\dagger$ is also a group; then $\mathfrak{M}$ is a group in $\mathfrak{G}$. If we let $\mathfrak{N}$ denote the intersection of all the algebraic sets in $\mathfrak{G}$ which contain $\mathfrak{M}$ then $\mathfrak{M} \subseteq \mathfrak{N} \subseteq \mathfrak{M}_0$. By Chapter I, Proposition 14, $\mathfrak{N}$ is an algebraic group. Also, the irreducible set $\mathfrak{M}_0$ is the union of $\mathfrak{N}$ and a finite union of irreducible sets of lower dimension than $\mathfrak{M}_0$, so that $\mathfrak{M}_0 = \mathfrak{N}$, $\mathfrak{M}_0$ is an algebraic group in $\mathfrak{G}$, and (Chapter II, Proposition 15) $\mathfrak{M}_0 = \mathfrak{M}$. Therefore $\mathfrak{M}$ is an irreducible algebraic group in $\mathfrak{G}$ of dimension

$$= \partial^0 \mathscr{G}\langle\rho\mathscr{G}\rangle/\mathscr{G} = \partial^0 \mathscr{G}_\rho/\mathscr{G} = \partial^0 \mathscr{G}^\dagger\langle\rho^\dagger\mathscr{G}^\dagger\rangle/\mathscr{G}^\dagger = \dim \mathfrak{M}^\dagger.$$

Thus every irreducible algebraic group in $\mathfrak{G}^\dagger$ is mapped onto an irreducible algebraic group in $\mathfrak{G}$ of the same dimension. Since every algebraic group is the union of the finite number of cosets of its component of the identity, a similar remark holds for not necessarily irreducible algebraic groups. In particular, the image of $\mathfrak{G}^\dagger$ is $\mathfrak{G}_1 = \mathfrak{G}(\mathscr{F}^\dagger \cap \mathscr{G})$.

COROLLARY. *Let $\mathscr{G}$ be strongly normal over $\mathscr{F}$, and let $\mathscr{G}_0, \mathscr{G}_1, \cdots, \mathscr{G}_r$ be differential fields such that*

$$\mathscr{F} = \mathscr{G}_0 \subseteq \mathscr{G}_1 \subseteq \cdots \subseteq \mathscr{G}_r,\ \mathscr{G} \subseteq \mathscr{G}_r,$$

and $\mathscr{G}_i$ is a strongly normal extension of $\mathscr{G}_{i-1}$ $(1 \leq i \leq r)$; denote the group of all automorphisms of $\mathscr{G}_i$ over $\mathscr{G}_{i-1}$ by $\mathfrak{G}_i$ $(1 \leq i \leq r)$. Then

$$\mathfrak{G} = \mathfrak{G}(\mathscr{G}_0 \cap \mathscr{G}) \supseteq \mathfrak{G}(\mathscr{G}_1 \cap \mathscr{G}) \supseteq \cdots \supseteq \mathfrak{G}(\mathscr{G}_r \cap \mathscr{G}) = \{\iota\} \tag{5}$$

is a normal chain of subgroups of $\mathfrak{G}$, $\mathfrak{G}_i(\mathscr{G}_i \cap \mathscr{G}\langle\mathscr{G}_{i-1}\rangle)$ is a normal subgroup of $\mathfrak{G}_i$, and for $i = 1, 2, \cdots, r$

$$\mathfrak{G}(\mathscr{G}_{i-1} \cap \mathscr{G})/\mathfrak{G}(\mathscr{G}_i \cap \mathscr{G}) \approx \mathfrak{G}_i/\mathfrak{G}_i(\mathscr{G}_i \cap \mathscr{G}\langle\mathscr{G}_{i-1}\rangle), \tag{6}$$

$$\begin{aligned} \dim \mathfrak{G}(\mathscr{G}_{i-1} \cap \mathscr{G}) &- \dim \mathfrak{G}(\mathscr{G}_i \cap \mathscr{G}) \\ &= \dim \mathfrak{G}_i - \dim \mathfrak{G}_i(\mathscr{G}_i \cap \mathscr{G}\langle\mathscr{G}_{i-1}\rangle). \end{aligned} \tag{7}$$

Proof. If $r=1$ the assertions follow immediately from Theorem 4. Let $r>1$ and suppose the corollary proved for lower values of r.

By the corollary to Theorem 4, $\mathcal{G}_1 \cap \mathcal{G}$ is strongly normal over $\mathcal{F}$ so that (Theorem 4) $\mathfrak{G}(\mathcal{G}_1 \cap \mathcal{G})$ is a normal subgroup $\mathfrak{G}=\mathfrak{G}(\mathcal{G}_0 \cap \mathcal{G})$ and $\mathfrak{G}_1(\mathcal{G}_1 \cap \mathcal{G})$ is a normal subgroup of $\mathfrak{G}_1$. The two factor groups $\mathfrak{G}/\mathfrak{G}(\mathcal{G}_1 \cap \mathcal{G})$ and $\mathfrak{G}_1/\mathfrak{G}_1(\mathcal{G}_1 \cap \mathcal{G})$ are isomorphic to the group of all automorphisms of $\mathcal{G}_1 \cap \mathcal{G}$ over $\mathcal{F}$, and therefore to each other, so that (6) holds for $i=1$. By Theorem 2, also (7) holds for $i=1$, as in that case both members equal $\partial^0(\mathcal{G}_1 \cap \mathcal{G})/\mathcal{F}$.

To complete the proof we consider the group $\mathfrak{G}^\dagger$ of all automorphisms of $\mathcal{G}\langle\mathcal{G}_1\rangle$ over $\mathcal{F}\langle\mathcal{G}_1\rangle=\mathcal{G}_1$. By the theorem $\mathcal{G}\langle\mathcal{G}_1\rangle$ is strongly normal over $\mathcal{G}_1$. By the corollary (case $r-1$)

$$
\begin{aligned}
(8)\qquad \mathfrak{G}^\dagger &= \mathfrak{G}^\dagger(\mathcal{G}_1 \cap \mathcal{G}\langle\mathcal{G}_1\rangle)\\
&\supseteq \mathfrak{G}^\dagger(\mathcal{G}_2 \cap \mathcal{G}\langle\mathcal{G}_1\rangle) \supseteq \cdots \supseteq \mathfrak{G}^\dagger(\mathcal{G}_r \cap \mathcal{G}\langle\mathcal{G}_1\rangle)=\{\iota\}
\end{aligned}
$$

is a normal chain of subgroups of $\mathfrak{G}^\dagger$, $\mathfrak{G}_i(\mathcal{G}_i \cap \mathcal{G}\langle\mathcal{G}_{i-1}\rangle)$ is a normal subgroup of $\mathfrak{G}_i$ $(2 \leq i \leq r)$, and

$$
\mathfrak{G}^\dagger(\mathcal{G}_{i-1} \cap \mathcal{G}\langle\mathcal{G}_1\rangle)/\mathfrak{G}^\dagger(\mathcal{G}_i \cap \mathcal{G}\langle\mathcal{G}_1\rangle) \approx \mathfrak{G}_i/\mathfrak{G}_i(\mathcal{G}_i \cap \mathcal{G}\langle\mathcal{G}_{i-1}\rangle),
$$

$$
\begin{aligned}
\dim \mathfrak{G}^\dagger(\mathcal{G}_{i-1} \cap \mathcal{G}\langle\mathcal{G}_1\rangle) &- \dim \mathfrak{G}^\dagger(\mathcal{G}_i \cap \mathcal{G}\langle\mathcal{G}_1\rangle)\\
&= \dim \mathfrak{G}_i - \dim \mathfrak{G}_i(\mathcal{G}_i \cap \mathcal{G}\langle\mathcal{G}_{i-1}\rangle)
\end{aligned}
$$

for $i=2,\cdots,r$. By the theorem, the mapping $\sigma^\dagger \rightarrow \sigma$ which assigns to each $\sigma^\dagger \varepsilon \mathfrak{G}^\dagger$ the restriction σ of $\sigma^\dagger$ to $\mathcal{G}$ is a dimension-preserving isomorphism which maps the normal chain (8) onto the normal chain

$$
\mathfrak{G}(\mathcal{G}_1 \cap \mathcal{G}) \supseteq \mathfrak{G}(\mathcal{G}_2 \cap \mathcal{G}) \supseteq \cdots \supseteq \mathfrak{G}(\mathcal{G}_r \cap \mathcal{G})=\{\iota\}.
$$

It now quickly follows that (6) and (7) hold for $2 \leq i \leq r$, and the proof of the corollary is complete.

4. Primitive elements. An element α will be called *primitive* over $\mathcal{F}$ if $\delta_i\alpha \,\varepsilon\, \mathcal{F}$ $(1 \leq i \leq m)$. It is obvious that if α is primitive over $\mathcal{F}$ with $\delta_i\alpha=a_i$ $(1 \leq i \leq m)$, and if β is primitive over $\mathcal{F}$ with $\delta_i\beta=b_i$ $(1 \leq i \leq m)$, and if we set $\eta=\alpha+\beta$, then η is primitive over $\mathcal{F}$ with $\delta_i\eta=a_i+b_i$ $(1 \leq i \leq m)$.

Let α be primitive over $\mathcal{F}$ and suppose that the field of constants of $\mathcal{F}\langle\alpha\rangle$ is $\mathcal{C}$. For every isomorphism σ of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$

$$
\delta_i(\sigma\alpha-\alpha)=\sigma\delta_i\alpha-\delta_i\alpha=\delta_i\alpha-\delta_i\alpha=0 \qquad (1 \leq i \leq m),
$$

so that $c(\sigma) = \sigma\alpha - \alpha$ is a constant. Because $\sigma\alpha = \alpha + c(\sigma)$ $\mathcal{F}\langle\alpha\rangle$ is strongly normal over $\mathcal{F}$ and every isomorphism of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ is strong. If σ_1, σ_2 are two isomorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ then

$$\sigma_1\sigma_2\alpha = \sigma_1(\alpha + c(\sigma_2)) = \alpha + c(\sigma_1) + c(\sigma_2),$$

so that $c(\sigma_1\sigma_2) = c(\sigma_1) + c(\sigma_2)$. Since $c(\sigma) = 0$ only when σ is the identity automorphism ι of $\mathcal{F}\langle\alpha\rangle$, it follows that $\sigma \to c(\sigma)$ is an isomorphism of the group of all automorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ into $\mathcal{C}^+$ (the additive group of $\mathcal{C}$). If $\alpha \in \mathcal{F}$ the automorphism group consists solely of ι, and the corresponding group in $\mathcal{C}^+$ is the zero group. Suppose $\alpha \notin \mathcal{F}$. Then there exists an automorphism σ of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ different from ι, and therefore with $c(\sigma) \neq 0$. Now if c is a constant then $\alpha + c$ is a generic specialization of α over $\mathcal{F}$ if and only if $\alpha + c$ is a specialization of α over $\mathcal{F}$, that is if and only if $\alpha + c$ is a zero of every differential polynomial in $\mathcal{F}\{y\}$ which vanishes at α, and this takes place if and only if c is a zero of a certain ideal of polynomials in $\mathcal{C}[u]$. Since this ideal of polynomials has infinitely many zeros, namely $c(\sigma^n) = nc(\sigma)$ for every integer n, it is the zero ideal, so that $\alpha + c$ is a generic specialization of α over $\mathcal{F}$ and $c = c(\sigma)$ for a suitable isomorphism σ of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ for every constant c. Summarizing: *Let α be primitive over $\mathcal{F}$ and the field of constants of $\mathcal{F}\langle\alpha\rangle$ be $\mathcal{C}$. Then $\mathcal{F}\langle\alpha\rangle$ is strongly normal over $\mathcal{F}$; either $\alpha \in \mathcal{F}$, or else α is transcendental over $\mathcal{F}$ and the mapping $\sigma \to \sigma\alpha - \alpha$ is an isomorphism of the group of all automorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ onto $\mathcal{C}^+$ and there exists no differential field between $\mathcal{F}$ and $\mathcal{F}\langle\alpha\rangle$ other than $\mathcal{F}$ and $\mathcal{F}\langle\alpha\rangle$.*

We note the trivial fact that, with the law of composition $(c_1, c_2) \to c_1 + c_2$, one-dimensional affine space becomes a group variety $\boldsymbol{D}$ in the sense of Weil [10], and that $\mathcal{C}^+$ is the subgroup of $\boldsymbol{D}$ consisting of all points of $\boldsymbol{D}$ which are rational over $\mathcal{C}$.

5. Exponential elements. An element α will be called *exponential* over $\mathcal{F}$ if $\alpha \neq 0$ and $\alpha^{-1}\delta_i\alpha \in \mathcal{F}$ $(1 \leq i \leq m)$. It is obvious that if α is exponential over $\mathcal{F}$ with $\alpha^{-1}\delta_i\alpha = a_i$ $(1 \leq i \leq m)$, and if β is exponential over $\mathcal{F}$ with $\beta^{-1}\delta_i\beta = b_i$ $(1 \leq i \leq m)$, and if we set $\eta = \alpha\beta$, the η is exponential over $\mathcal{F}$ with $\eta^{-1}\delta_i\eta = a_i + b_i$ $(1 \leq i \leq m)$.

Let α be exponential over $\mathcal{F}$ and suppose that the field of constants of $\mathcal{F}\langle\alpha\rangle$ is $\mathcal{C}$. For every isomorphism σ of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$

$$\begin{aligned}\delta_i(\alpha^{-1}\sigma\alpha) &= \alpha^{-2}(\alpha \cdot \delta_i\sigma\alpha - \sigma\alpha \cdot \delta_i\alpha) \\ &= \alpha^{-2}(\alpha \cdot \sigma\alpha \cdot \sigma(\alpha^{-1}\delta_i\alpha) - \sigma\alpha \cdot \alpha \cdot \alpha^{-1}\delta_i\alpha) = 0 \qquad (1 \leq i \leq m),\end{aligned}$$

so that $c(\sigma) = \alpha^{-1}\sigma\alpha$ is a constant. Because $\sigma\alpha = c(\sigma)\alpha$, $\mathcal{F}\langle\alpha\rangle$ is strongly normal over $\mathcal{F}$ and every isomorphism of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ is strong. If σ_1, σ_2 are two isomorphism of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ then

$$\sigma_1\sigma_2\alpha = \sigma_1(c(\sigma_2)\alpha) = c(\sigma_1)c(\sigma_2)\alpha,$$

so that $c(\sigma_1\sigma_2) = c(\sigma_1)c(\sigma_2)$. Since $c(\sigma) = 1$ only when $\sigma = \iota$, it follows that $\sigma \to c(\sigma)$ is an isomorphism of the group of all automorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ into $\mathcal{C}^\times$ (the multiplicative group of nonzero elements of $\mathcal{C}$). If α is algebraic over $\mathcal{F}$ then the automorphism group is finite, say of order d; since every finite subgroup of $\mathcal{C}^\times$ is cyclic the automorphism group is cyclic, being generated by a single automorphism, say σ. It follows that $c(\sigma)$ is a primitive d-th root of unity, so that $\sigma(\alpha^d) = (c(\sigma)\alpha)^d = \alpha^d$, whence $\alpha^d \in \mathcal{F}$. If α is transcendental over $\mathcal{F}$ then the automorphism group is infinite, and it follows, much as in the case of primitive elements, that $c\alpha$ is a generic specialization of α over $\mathcal{F}$ for every nonzero constant c, so that $\sigma \to c(\sigma)$ is an isomorphism of the automorphism group of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ onto $\mathcal{C}^\times$; furthermore, if $\mathcal{F}_1$ is a differential field between $\mathcal{F}$ and $\mathcal{F}\langle\alpha\rangle$ then, since α is exponential over $\mathcal{F}_1$, the mapping $\sigma \to c(\sigma)$ maps the group of all automorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}_1$ either onto a cyclic group of some finite order d (in which case $\alpha^d \in \mathcal{F}_1$, whence $\mathcal{F}_1 = \mathcal{F}\langle\alpha^d\rangle$) or else onto the whole group $\mathcal{C}^\times$ (in which case $\mathcal{F}_1 = \mathcal{F}$). Summarizing: *Let α be exponential over $\mathcal{F}$ and the field of constants of $\mathcal{F}\langle\alpha\rangle$ be $\mathcal{C}$. Then $\mathcal{F}\langle\alpha\rangle$ is strongly normal over $\mathcal{F}$; either there exists an integer $d > 0$ such that $\alpha^d \in \mathcal{F}$, or else α is transcendental over $\mathcal{F}$ and the mapping $\sigma \to \alpha^{-1}\sigma\alpha$ is an isomorphism of the group of all automorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ onto $\mathcal{C}^\times$ and the only differential fields between $\mathcal{F}$ and $\mathcal{F}\langle\alpha\rangle$ are those of the form $\mathcal{F}\langle\alpha^d\rangle$, where d is an integer ≥ 0.*

We note the trivial fact that, with the law of composition $(c_1, c_2) \to c_1c_2$, one-dimensional affine space with the origin deleted becomes a group variety $\boldsymbol{E}$, and that $\mathcal{C}^\times$ is the subgroup of $\boldsymbol{E}$ consisting of all points of $\boldsymbol{E}$ which are rational over $\mathcal{C}$.

6. Weierstrassian elements. An element α will be called *weierstrassian* over $\mathcal{F}$ if α is not a constant and there exist two elements $g_2, g_3 \in \mathcal{C}$ with $27g_3^2 - g_2^3 \neq 0$ and m elements $a_1, \cdots, a_m \in \mathcal{F}$ such that

$$(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3) \qquad (1 \leq i \leq m).$$

The condition $27g_3^2 - g_2^3 \neq 0$ is equivalent to the condition that the polynomial $4y^3 - g_2y - g_3$ have no multiple root; the constants g_2, g_3, which are

uniquely determined if α is transcendental over $\mathcal{F}$, will be called the *invariants* of α. Since any a_i may obviously be replaced by $-a_i$ *we may suppose (and in the future we always shall suppose, without expressly mentioning the fact) that* $(\delta_1\alpha:\cdots:\delta_m\alpha)=(a_1:\cdots:a_m)$.

In order to study weierstrassian elements with invariants g_2, g_3 we consider the irreducible algebraic curve in the projective plane defined by the equation

$$X_0X_2^2-4X_1^3+g_2X_0^2X_1+g_3X_0^3=0.$$

This curve, which is of genus 1, has precisely one point on the line at infinity $X_0=0$, namely the point $(0:0:1)$. On this curve there exists a law of composition, which we shall write multiplicatively, with respect to which the points of the curve form a group, the unity element being the point of infinity; the curve is, in the language of Weil [10], a group variety (actually an abelian variety) of dimension 1. If $(1:\xi_1:\xi_2)$ and $(1:\eta_1:\eta_2)$ are two points of this curve and if $\eta_1\neq\xi_1$, then their product is given by the formulae

$$(9)\quad\begin{cases}(1:\xi_1:\xi_2)(1:\eta_1:\eta_2)=(1:\zeta_1:\zeta_2),\\ \zeta_1=-\xi_1-\eta_1+\frac{1}{4}\left(\dfrac{\xi_2-\eta_2}{\xi_1-\eta_1}\right)^2,\\ \zeta_2=-\frac{1}{2}(\xi_2+\eta_2)+\frac{3}{2}(\xi_1+\eta_1)\dfrac{\xi_2-\eta_2}{\xi_1-\eta_1}-\frac{1}{4}\left(\dfrac{\xi_2-\eta_2}{\xi_1-\eta_1}\right)^3;\end{cases}$$

the inverse of any point $(1:\xi_1:\xi_2)$ of the curve is the point $(1:\xi:-\xi_2)$, so that there are precisely three points of order 2, namely the points $(1:e_1:0)$, $(1:e_2:0)$, $(1:e_3:0)$, where e_1, e_2, e_3 are the roots of $4y^3-g_2y-g_3$; the square of any point $(1:\xi_1:\xi_2)$ with $\xi_2\neq0$ is given by the formulae

$$(10)\quad\begin{cases}(1:\xi_1:\xi_2)^2=(1:\zeta_1:\zeta_2),\\ \zeta_1=-2\xi_1+4^{-1}(6\xi_1^2-\frac{1}{2}g_2)^2\xi_2^{-2},\\ \zeta_2=-\xi_2+3\xi_1(6\xi_1^2-\frac{1}{2}g_2)\xi_2^{-1}-4^{-1}(6\xi_1^2-\frac{1}{2}g_2)^3\xi_2^{-3}.\end{cases}$$

These facts are well-known and are not difficult to verify directly. We shall denote this group variety by $\mathbf{W}(g_2, g_3)$.

If α is weierstrassian over $\mathcal{F}$ with $(\delta_i\alpha)^2=a_i^2(4\alpha^3-g_2\alpha-g_3)$ $(1\leq i\leq m)$ then, since α is not a constant, $\delta_i\alpha\neq0$ for at least one value of i and $a_i\neq0$ whenever $\delta_i\alpha\neq0$. By the convention made above, the point $(1:\alpha:a_i^{-1}\delta_i\alpha)$ does not depend on the choice of i from among those values for which $a_i\neq0$; this point obviously belongs to $\mathbf{W}(g_2, g_3)$.

LEMMA 2. *Let α be weierstrassian over $\mathcal{F}$ with*

$$(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3) \qquad (1 \leq i \leq m)$$

and $a_{i_0} \neq 0$, let β be weierstrassian over $\mathcal{F}$ with

$$(\delta_i\beta)^2 = b_i^2(4\beta^3 - g_2\beta - g_3) \qquad (1 \leq i \leq m)$$

and $b_{j_0} \neq 0$, suppose that

$$(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)(1 : \beta : b_{j_0}^{-1}\delta_{j_0}\beta) \neq (0 : 0 : 1),$$

and let

$$(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)(1 : \beta : b_{j_0}^{-1}\delta_{j_0}\beta) = (1 : \eta : \zeta).$$

Then $\delta_i\eta = (a_i + b_i)\zeta$ $(1 \leq i \leq m)$, so that either η and ζ are both constants or else η is weierstrassian over $\mathcal{F}$ with

$$(\delta_i\eta)^2 = (a_i + b_i)^2(4\eta^3 - g_2\eta - g_3) \qquad (1 \leq i \leq m).$$

Proof. Suppose first that $\alpha \neq \beta$. Then by (9) and a simple computation we find that

$$\begin{aligned}(11) \quad \zeta = {} & (\alpha - \beta)^{-3}(-\beta^2(3\alpha + \beta) + \tfrac{1}{4}g_2(\alpha + 3\beta) + g_3)a_{i_0}^{-1}\delta_{i_0}\alpha \\ & + (\alpha - \beta)^{-3}(\alpha^2(\alpha + 3\beta) - \tfrac{1}{4}g_2(3\alpha + \beta) - g_3)b_{j_0}^{-1}\delta_{j_0}\beta.\end{aligned}$$

On the other hand

$$\delta_i\alpha = a_i a_{i_0}^{-1}\delta_{i_0}\alpha, \qquad \delta_i\beta = b_i b_{j_0}^{-1}\delta_{j_0}\beta,$$

$$\delta_i(a_{i_0}^{-1}\delta_{i_0}\alpha) = (6\alpha^2 - \tfrac{1}{2}g_2)a_i, \qquad \delta_i(b_{j_0}^{-1}\delta_{j_0}\beta) = (6\beta^2 - \tfrac{1}{2}g_2)b_i,$$

so that from (9) we find that

$$\begin{aligned}\delta_i\eta = {} & -\delta_i\alpha - \delta_i\beta + \tfrac{1}{2}\,\frac{a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta}{\alpha - \beta}\left(\frac{\delta_i(a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta)}{\alpha - \beta}\right. \\ & \left. - \frac{(a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}\delta_{j_0}\beta)(\delta_i\alpha - \delta_i\beta)}{(\alpha - \beta)^2}\right) \\ = {} & - a_i a_{i_0}^{-1}\delta_{i_0}\alpha - b_i b_{j_0}^{-1}\delta_{j_0}\beta \\ & + \tfrac{1}{2}\,\frac{a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta}{\alpha - \beta}\;\frac{(6\alpha^2 - \tfrac{1}{2}g_2)a_i - (6\beta^2 - \tfrac{1}{2}g_2)b_i}{\alpha - \beta} \\ & - \tfrac{1}{2}\,\frac{(a_{i_0}^{-1}\delta_{i_0}\alpha - b_{j_0}^{-1}\delta_{j_0}\beta)^2(a_i a_{i_0}^{-1}\delta_{i_0}\alpha - b_i b_{j_0}^{-1}\delta_{j_0}\beta)}{(\alpha - \beta)^3}.\end{aligned}$$

The coefficient of a_i here is easily seen to be the second member of (11), and likewise for the coefficient of b_i here. It follows that $\delta_i\eta = (a_i + b_i)\zeta$.

Now suppose that $\alpha = \beta$. Because by hypothesis

$$(1 : \alpha : a_{i_0}^{-1}\delta_{i_0}\alpha)(1 : \beta : b_{j_0}^{-1}\delta_{j_0}\beta) \neq (0 : 0 : 1),$$

we have $a_i = b_i$ $(1 \leq i \leq m)$. Therefore (10) is applicable in computing $(1:\eta:\zeta)$, and we find on the one hand

$$\begin{aligned}\zeta &= -a_{i_0}^{-1}\delta_{i_0}\alpha + 3\alpha(6\alpha^2 - \tfrac{1}{2}g_2)(a_{i_0}^{-1}\delta_{i_0}\alpha)^{-1} \\ &\quad - \tfrac{1}{4}(6\alpha^2 - \tfrac{1}{2}g_2)^3(a_{i_0}^{-1}\delta_{i_0}\alpha)^{-3} \\ &= (-1 + 3\alpha(6\alpha^2 - \tfrac{1}{2}g_2)(4\alpha^3 - g_2\alpha - g_3)^{-1} \\ &\quad - \tfrac{1}{4}(6\alpha^2 - \tfrac{1}{2}g_2)^3(4\alpha^3 - g_2\alpha - g_3)^{-2})a_{i_0}^{-1}\delta_{i_0}\alpha,\end{aligned}$$

and on the other hand

$$\eta = -2\alpha + \tfrac{1}{4}(6\alpha^2 - \tfrac{1}{2}g_2)^2(4\alpha^3 - g_2\alpha - g_3)^{-1},$$

so that

$$\begin{aligned}\delta_i\eta &= (-2 + 6\alpha(6\alpha^2 - \tfrac{1}{2}g_2)(4\alpha^3 - g_2\alpha - g_3)^{-1} \\ &\quad - \tfrac{1}{2}(6\alpha^2 - \tfrac{1}{2}g_2)^3(4\alpha^3 - g_2\alpha - g_3)^{-2})\delta_i\alpha,\end{aligned}$$

whence $\delta_i\eta = 2a_i\zeta = (a_i + b_i)\zeta$, q. e. d.

Now let α be an element which is weierstrassian over $\mathcal{F}$ with

$$(\delta_i\alpha)^2 = a_i^2(4\alpha^3 - g_2\alpha - g_3) \qquad (1 \leq i \leq m),$$

and suppose that the field of constants of $\mathcal{F}\langle\alpha\rangle$ is $\mathcal{C}$. We let i_0 denote any subscript such that $a_{i_0} \neq 0$. For any isomorphism σ of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$,

$$\begin{aligned}&(1:\sigma\alpha:a_{i_0}^{-1}\delta_{i_0}\sigma\alpha)(1:\alpha:-a_{i_0}^{-1}\delta_{i_0}\alpha) \\ &\qquad = (1:\sigma\alpha:a_{i_0}^{-1}\delta_{i_0}\sigma\alpha)(1:\alpha:a_{i_0}^{-1}\delta_{i_0}\alpha)^{-1}\end{aligned}$$

is a point of the group variety $\boldsymbol{W}(g_2, g_3)$; we denote this point by $P(\sigma)$. If $\sigma = \iota$ then obviously $P(\sigma) = (0:0:1)$. Suppose $\sigma \neq \iota$; then $P(\sigma) \neq (0:0:1)$, and we may write $P(\sigma) = (1:c_1(\sigma):c_2(\sigma))$. It follows from Lemma 2 that $\delta_i c_1(\sigma) = (a_i - a_i)c_2(\sigma) = 0$ $(1 \leq i \leq m)$, so that $c_1(\sigma)$ and, therefore, $c_2(\sigma)$ are constants. Therefore for every isomorphism σ of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ we have

$$(1:\sigma\alpha:a_{i_0}^{-1}\delta_{i_0}\sigma\alpha) = P(\sigma)(1:\alpha:a_{i_0}^{-1}\delta_{i_0}\alpha)$$

and $\sigma\alpha \in \mathcal{F}\langle\alpha\rangle\langle\mathcal{C}^*\rangle$, so that $\mathcal{F}\langle\alpha\rangle$ is strongly normal over $\mathcal{F}$ and σ is strong. If σ_1, σ_2 are two isomorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ then

$$\begin{aligned}&(1:\sigma_1\sigma_2\alpha:a_{i_0}^{-1}\delta_{i_0}\sigma_1\sigma_2\alpha) \\ &\qquad = P(\sigma_2)(1:\sigma_1\alpha:c_{i_0}^{-1}\delta_{i_0}\sigma_1\alpha) = P(\sigma_2)P(\sigma_1)(1:\alpha:a_{i_0}^{-1}\delta_{i_0}\alpha),\end{aligned}$$

so that $P(\sigma_1\sigma_2) = P(\sigma_1)P(\sigma_2)$. Since $P(\sigma) = (0:0:1)$ only when $\sigma = \iota$, it follows that $\sigma \to P(\sigma)$ is an isomorphism of the group of all automorphisms

of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ into the group $W(g_2, g_3; \mathcal{C})$ consisting of all points of $W(g_2, g_3)$ which are rational over $\mathcal{C}$.

Let c_1, c_2 be any two constants such that $(1:c_1:c_2) \varepsilon W(g_2, g_3)$ and set $(1:\beta:\beta') = (1:c_1:c_2)(1:\alpha:a_{i_0}{}^{-1}\delta_{i_0}\alpha)$. Now if (β, β') is a specialization of $(\alpha, a_{i_0}{}^{-1}\delta_{i_0}\alpha)$ over $\mathcal{F}$ then the specialization is generic, and this will be the case if and only if $\beta = \sigma\alpha$, $\beta' = a_{i_0}{}^{-1}\delta_{i_0}\sigma\alpha$ for some isomorphism σ of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$, that is, if and only if $(1:c_1:c_2) = P(\sigma)$ for some such σ. On the other hand (β, β') is a specialization of $(\alpha, a_{i_0}{}^{-1}\delta_{i_0}\alpha)$ over $\mathcal{F}$ if and only if (c_1, c_2) is a zero of a certain set of polynomials with coefficients in $\mathcal{F}\langle\alpha\rangle$, and therefore if and only if (c_1, c_2) is a zero of a certain set of polynomials with coefficients in $\mathcal{C}$. It follows that $\sigma \to P(\sigma)$ maps the group of all automorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ onto the intersection with $W(g_2, g_3; \mathcal{C})$ of a subgroup of $W(g_2, g_3)$ which is a subvariety (not necessarily irreducible) of $W(g_2, g_3)$. Of course the subvarieties of an irreducible curve other than the curve itself are finite.

We may summarize these facts as follows: *Let α be weierstrassian over $\mathcal{F}$ and the field of constants of $\mathcal{F}\langle\alpha\rangle$ be $\mathcal{C}$. Then $\mathcal{F}\langle\alpha\rangle$ is strongly normal over $\mathcal{F}$; either α is algebraic over $\mathcal{F}$ and the mapping*

$$\sigma \to (1:\alpha:a_{i_0}{}^{-1}\delta_{i_0}\alpha)^{-1}(1:\sigma\alpha:a_{i_0}{}^{-1}\delta_{i_0}\sigma\alpha)$$

is an isomorphism of the group of all automorphisms of $\mathcal{F}\langle\alpha\rangle$ over $\mathcal{F}$ onto a finite subgroup of $W(g_2, g_3; \mathcal{C})$, or else α is transcendental over $\mathcal{F}$ and this mapping is an isomorphism onto $W(g_2, g_3; \mathcal{C})$. It can be shown, although we do not do so here, that if $\mathcal{F}_1$ is a differential field between $\mathcal{F}$ and $\mathcal{F}\langle\alpha\rangle$ other than $\mathcal{F}$ then $\mathcal{F}_1$ contains an element α_1 weierstrassian over $\mathcal{F}$ such that $\mathcal{F}_1 = \mathcal{F}\langle\alpha_1\rangle$.

7. Picard-Vessiot extensions. Let $\mathcal{G}$ be a Picard-Vessiot extension of $\mathcal{F}$. Then (Kolchin [3] and [6]), for suitable generators $\eta_1, \cdots, \eta_n$ of $\mathcal{G}$ over $\mathcal{F}$, every isomorphism σ of $\mathcal{G}$ over $\mathcal{F}$ satisfies equations

$$\sigma\eta_j = \sum_{i=1}^{n} c_{ij}(\sigma)\eta_i \qquad (1 \leq j \leq n), \tag{12}$$

where each $c_{ij}(\sigma)$ is a constant, and these equations establish a one-to-one correspondence between the set of all isomorphisms σ of $\mathcal{G}$ over $\mathcal{F}$ and a certain set of invertible matrices $(c_{ij}(\sigma))$ of degree n with constant coordinates. It follows that $\mathcal{G}$ is strongly normal over $\mathcal{F}$, and each isomorphism of $\mathcal{G}$ over $\mathcal{F}$ is strong. The mapping $\sigma \to (c_{ij}(\sigma))$ is an isomorphism of the algebraic group $\mathfrak{G}$ of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ onto a certain algebraic

matric group $\mathfrak{G}_M$ over $\mathcal{C}$; algebraic subgroups of $\mathfrak{G}$ are mapped thereby onto algebraic subgroups of $\mathfrak{G}_M$ of the same dimension. If σ, τ are isomorphisms of $\mathcal{G}$ over $\mathcal{F}$ then τ is a specialization of σ if and only if $(c_{ij}(\tau))$ is a specialization of $(c_{ij}(\sigma))$ over $\mathcal{C}$.

We shall say that a differential field $\mathcal{H}$ is an extension of $\mathcal{F}$ by algebraic, primitive, exponential and weierstrassian elements if $\mathcal{H}$ contains a finite family of elements $\alpha_1, \cdots, \alpha_r$ such that $\mathcal{H} = \mathcal{F}\langle\alpha_1, \cdots, \alpha_r\rangle$ and for each i $(1 \leq i \leq r)$ α_i is either algebraic, or primitive, or exponential, or weierstrassian over $\mathcal{F}\langle\alpha_1, \cdots, \alpha_{i-1}\rangle$. If $\mathcal{H} = \mathcal{F}\langle\alpha_1, \cdots, \alpha_r\rangle$ and for each i $(1 \leq i \leq r)$ α_i is either algebraic, or primitive, or exponential over $\mathcal{F}\langle\alpha_1, \cdots, \alpha_{i-1}\rangle$, and if the field of constants of $\mathcal{H}$ is $\mathcal{C}$, then $\mathcal{H}$ is called a *liouvillian* extension of $\mathcal{F}$.

THEOREM 6. *If a Picard-Vessiot extension of $\mathcal{F}$ is contained in an extension of $\mathcal{F}$ by algebraic, primitive, exponential, and weierstrassian elements with field of constants $\mathcal{C}$,*[8] *then the Picard-Vessiot extension is a liouvillian extension of $\mathcal{F}$.*

Proof. By the hypothesis, the results of the preceding three sections, and the corollary to Theorem 5, the group of all automorphisms of the Picard-Vessiot extension of $\mathcal{F}$ has a normal chain of algebraic subgroups in which each factor group is either finite or abelian. Therefore (Kolchin [3], § 8, Theorem 1) the component of the identity of this group of automorphisms is solvable. It follows that the Picard-Vessiot extension is a liouvillian extension.

8. Extensions of transcendence degree 1; formulation of the theorem. In §§ 4-6 we saw that if α is transcendental and either primitive or exponential or weierstrassian over $\mathcal{F}$, and if the field of constants of $\mathcal{F}\langle\alpha\rangle$ is $\mathcal{C}$, then $\mathcal{F}\langle\alpha\rangle$ is a strongly normal extension of $\mathcal{F}$ of transcendence degree 1. We shall now state a theorem which implies that every strongly normal (indeed, every weakly normal) extension of $\mathcal{F}$ of transcendence degree 1 can be obtained by combining the adjunction of an element of one of these three types with algebraic adjunctions.

Let $\mathcal{G}$ be a weakly normal extension of $\mathcal{F}$. It is a simple matter to see that the relative algebraic closure $\mathcal{F}^0$ of $\mathcal{F}$ in $\mathcal{G}$ is a normal algebraic extension of $\mathcal{F}$ (in the classical sense) of finite degree. If the group $\mathfrak{G}$ of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ is finite then $\mathcal{G} = \mathcal{F}^0$; therefore if $\mathcal{G}$ is trans-

[8] By considerations similar to those of Kolchin [4] it can be shown that this restriction on the field of constants of the extension may be omitted.

cendental over $\mathfrak{F}$ then $\mathfrak{G}$ is infinite, and it easily follows that the group $\mathfrak{G}^0$ of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}^0$ is infinite. We therefore may state our theorem in the following form.

THEOREM 7. *Let $\mathfrak{G}$ be of transcendence degree* 1 *over $\mathfrak{F}$, let $\mathfrak{F}$ be relatively algebraically closed in $\mathfrak{G}$, and let the group $\mathfrak{G}$ of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ be infinite. Then there exists an element $\alpha \varepsilon \mathfrak{G}$ such that either α is primitive over $\mathfrak{F}$ and $\mathfrak{G} = \mathfrak{F}\langle\alpha\rangle$, or α is exponential over $\mathfrak{F}$ and $\mathfrak{G} = \mathfrak{F}\langle\alpha\rangle$, or α is weierstrassian over $\mathfrak{F}$ and $\mathfrak{G}$ is an algebraic*[9] *extension of $\mathfrak{F}\langle\alpha\rangle$. In the last case, if $\mathfrak{F}$ is algebraically closed then the weierstrassian element α may be chosen so that $\mathfrak{G} = \mathfrak{F}\langle\alpha\rangle$.*

This theorem will be proved in §§ 9-11.

An immediate consequence of Theorem 7, the results of §§ 4-6, and the Corollary of Theorem 5, is the following.

COROLLARY. *Let $\mathfrak{G}$ be any strongly normal extension of $\mathfrak{F}$. If $\mathfrak{G}$ is contained in an extension of $\mathfrak{F}$ by algebraic, primitive, exponential, and weierstrassian elements then the group $\mathfrak{G}$ of all automorphisms of $\mathfrak{G}$ over $\mathfrak{F}$ has a normal chain $\mathfrak{G} = \mathfrak{G}_0 \supseteq \cdots \supseteq \mathfrak{G}_s = \{\iota\}$ of algebraic groups such that* $\dim \mathfrak{G}_{i-1} - \dim \mathfrak{G}_i \leq 1$ $(1 \leq i \leq s)$. *Conversely, if $\mathfrak{G}$ has such a normal chain then $\mathfrak{G}$ is itself an extension of $\mathfrak{F}$ by algebraic, primitive, exponential, and weierstrassian elements.*

9. The proof begun: reduction to the case of algebraically closed ground field. We shall show in this section that if Theorem 7 holds when $\mathfrak{F}$ is algebraically closed then it holds in general.

Let $\mathfrak{F}^\dagger$ be the algebraic closure of $\mathfrak{F}$ and let $\mathfrak{G}^\dagger = \mathfrak{G}\langle\mathfrak{F}^\dagger\rangle$. Since $\mathfrak{F}$ is relatively algebraically closed in $\mathfrak{G}$, the degree of each element of $\mathfrak{F}^\dagger$ over $\mathfrak{F}$ equals its degree over $\mathfrak{G}$. Now every $P \varepsilon \mathfrak{F}^\dagger\{y_1, \cdots, y_n\}$ can be written in the form $P = \sum_{i=0}^{d-1} P_i\lambda^i$, where each $P_i \varepsilon \mathfrak{F}\{y_1, \cdots, y_n\}$ and λ is an element of $\mathfrak{F}^\dagger$ of some degree d over $\mathfrak{F}$, and therefore over $\mathfrak{G}$. Consequently a family $(\eta_1, \cdots, \eta_n)$ of elements of $\mathfrak{G}$ is a zero of P if and only if it is a zero of $P_0, P_1, \cdots, P_{d-1}$. From this it follows that every automorphism of $\mathfrak{G}$ over $\mathfrak{F}$ can be extended to an automorphism of $\mathfrak{G}^\dagger$ over $\mathfrak{F}^\dagger$. From the hypothesis of Theorem 7 it therefore follows that there are infinitely many automorphisms of $\mathfrak{G}^\dagger$ over $\mathfrak{F}^\dagger$. By the assumption that the theorem holds when the ground

[9] Abelian.

field is algebraically closed we conclude that there exists an element η such that $\mathcal{G}^\dagger = \mathcal{F}^\dagger\langle\eta\rangle$ and η is either primitive or exponential or weierstrassian over $\mathcal{F}^\dagger$. Thus there exist elements $a_1, \cdots, a_m \varepsilon \mathcal{F}^\dagger$ such that either $\delta_i\eta = a_i$ $(1 \leq i \leq m)$ or $\eta^{-1}\delta_i\eta = a_i$ $(1 \leq i \leq m)$ or $(4\eta^3 - g_2\eta - g_3)^{-1}(\delta_i\eta)^2 = a_i^2$ $(1 \leq i \leq m)$, where in the last case $g_2, g_3 \varepsilon \mathcal{C}$, $27g_3^2 - g_2^3 \neq 0$, and $(\delta_1\eta : \cdots : \delta_m\eta) = (a_1 : \cdots : a_m)$.

Let σ be any automorphism of $\mathcal{G}^\dagger$ over $\mathcal{F}^\dagger$ other than the identity. We may write, in the respective cases, $\sigma\eta = \eta + c$ or $\sigma\eta = c\eta$ or $(1 : \sigma\eta : a_i^{-1}\delta_i\sigma\eta) = (1 : c_1 : c_2)(1 : \eta : a_i^{-1}\delta_i\eta)$, where in the first case $c \varepsilon \mathcal{C}^+$, in the second case $c \varepsilon \mathcal{C}^\times$, and in the third case $(1 : c_1 : c_2) \varepsilon W(g_2, g_3 ; \mathcal{C})$ and $a_i \neq 0$. Now let τ be any automorphism of $\mathcal{G}^\dagger$ over $\mathcal{G}$. It is clear that $\tau\mathcal{F}^\dagger = \mathcal{F}^\dagger$; therefore $\sigma\tau\theta = \tau\theta = \tau\sigma\theta$ for all $\theta \varepsilon \mathcal{F}^\dagger$. If σ happens to be one of the infinitely many automorphisms of $\mathcal{G}^\dagger$ over $\mathcal{F}^\dagger$ which are extensions of automorphisms of $\mathcal{G}$ over $\mathcal{F}$ then $\sigma\tau\theta = \sigma\theta = \tau\sigma\theta$ for every $\theta \varepsilon \mathcal{G}$ whence, since $\mathcal{G}^\dagger = \mathcal{G}\langle\mathcal{F}^\dagger\rangle$, $\sigma\tau = \tau\sigma$. Because $\mathcal{G}^\dagger = \mathcal{F}^\dagger\langle\eta\rangle$ we may write $\tau\eta = f(\eta)$, where $f \varepsilon \mathcal{F}^\dagger\langle y\rangle$, and for arbitrary σ we shall have $\sigma\tau = \tau\sigma$ if and only if $\sigma\tau\eta = \tau\sigma\eta$, that is $f(\eta + c) = f(\eta) + c$ in the first case, $f(c\eta) = cf(\eta)$ in the second case, and

$$f\left(-\eta - c_1 + \tfrac{1}{4}\left(\frac{a_i^{-1}\delta_i\eta - c_2}{\eta - c_1}\right)^2\right) = -f(\eta) - c_1 + \tfrac{1}{4}\left(\frac{a_i^{-1}\delta_i f(\eta) - c_2}{f(\eta) - c_1}\right)^2$$

in the third case. Since this condition is satisfied for infinitely many choices of $c \varepsilon \mathcal{C}^+$ in the first case, of $c \varepsilon \mathcal{C}^\times$ in the second case, and of $(1 : c_1 : c_2) \varepsilon W(g_2, g_3 ; \mathcal{C})$ in the third case, it must be satisfied identically. Thus σ commutes with every automorphism of $\mathcal{G}^\dagger$ over $\mathcal{G}$.

Let $\tau_1, \cdots, \tau_n$ be automorphisms of $\mathcal{G}^\dagger$ over $\mathcal{G}$ such that the restrictions of $\tau_1, \cdots, \tau_n$ to $\mathcal{G}\langle\eta\rangle$ are distinct and constitute the set of all isomorphisms of $\mathcal{G}\langle\eta\rangle$ over $\mathcal{G}$ (so that n equals the degree of $\mathcal{G}\langle\eta\rangle$ over $\mathcal{G}$). Since $\sigma\tau_j = \tau_j\sigma$, we have, in the respective cases, $\sigma\tau_j\eta = \tau_j\eta + c$, or $\sigma\tau_j\eta = c\tau_j\eta$, or

$$(1 : \sigma\tau_j\eta : \sigma\tau_j(a_i^{-1}\delta_i\eta)) = (1 : c_1 : c_2)(1 : \tau_j\eta : \tau_j(a_i^{-1}\delta_i\eta)).$$

We now consider the first case. Letting $\alpha = \sum_j \tau_j\eta$, we see that $\alpha \varepsilon \mathcal{G}$; also, $\delta_i\alpha = \sum_j \tau_j a_i \varepsilon \mathcal{F}$, so that α is primitive over $\mathcal{F}$. Furthermore, $\sigma\alpha = \alpha + nc$, so that $\alpha \not\varepsilon \mathcal{F}^\dagger$; it follows (§ 4) that $\mathcal{G}^\dagger = \mathcal{F}^\dagger\langle\eta\rangle = \mathcal{F}^\dagger\langle\alpha\rangle$. If θ is any element of $\mathcal{G}$ then we may write $\theta = \sum \phi_i\alpha^i / \sum \psi_i\alpha^i$, where $\sum \phi_i y^i$ and $\sum \psi_i y^i$ are relatively prime polynomials in $\mathcal{F}^\dagger[y]$ such that the leading coefficient in $\sum \psi_i y^i$ is 1. For any automorphism τ of $\mathcal{G}^\dagger$ over $\mathcal{G}$ we have $\tau\alpha = \alpha$ and $\tau\theta = \theta$, so that $\sum \tau\phi_i\alpha^i / \sum \tau\psi_i\alpha^i = \sum \phi_i\alpha^i / \sum \psi_i\alpha^i$, whence $(\sum \tau\phi_i\alpha^i)(\sum \psi_i\alpha^i) = (\sum \phi_i\alpha^i)(\sum \tau\psi_i\alpha^i)$. Because of the relative primeness

mentioned above and the fact that $\sum \psi_i y^i$ has leading coefficient 1 we conclude that $\sum \tau\phi_i\alpha^i = \sum \phi_i\alpha^i$ and $\sum \tau\psi_i\alpha^i = \sum \psi_i\alpha^i$, so that $\tau\phi_i = \phi_i$ and $\tau\psi_i = \psi_i$ for all i. Since τ is any automorphism of $\mathcal{G}^\dagger$ over $\mathcal{G}$ it follows that ϕ_i and ψ_i belong to $\mathcal{G} \cap \mathcal{F}^\dagger = \mathcal{F}$, so that $\theta \varepsilon \mathcal{F}\langle\alpha\rangle$. Thus $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$, and the reduction is complete in the first case.

We turn to the second case. We assert that there exist an integer $r > 0$ and a nonzero element $\phi \varepsilon \mathcal{F}^\dagger$ such that $(\phi\eta)^r \varepsilon \mathcal{G}$. Indeed, $\sigma \prod \tau_j\eta = c^n \prod \tau_j\eta$, so that if we set $\chi = \eta^{-n} \prod \tau_j\eta$ then $\sigma\chi = \chi$, whence $\chi \varepsilon \mathcal{F}^\dagger$; letting ψ be an element of $\mathcal{F}^\dagger$ such that $\psi^n = \chi$, we find that $(\psi\eta)^n = \prod \tau_j\eta \varepsilon \mathcal{G}$, which proves our assertion. Of all pairs r, ϕ as above let us suppose we have chosen one for which r is as small as possible. Let $\alpha = (\phi\eta)^r$, so that $\alpha \varepsilon \mathcal{G}$; α is exponential over $\mathcal{F}^\dagger$, so that $\alpha^{-1}\delta_i\alpha \varepsilon \mathcal{F}^\dagger \cap \mathcal{G} = \mathcal{F}$ $(1 \leq i \leq m)$, whence α is exponential over $\mathcal{F}$. We shall show that $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$, thereby completing the reduction in the second case. Indeed, if θ is any element of $\mathcal{G}$ then $\theta \varepsilon \mathcal{G}^\dagger = \mathcal{F}^\dagger\langle\eta\rangle = \mathcal{F}^\dagger\langle\phi\eta\rangle$, so that we may write $\theta = \sum \phi_i(\phi\eta)^i / \sum \psi_i(\phi\eta)^i$, where $\sum \phi_i y^i$ and $\sum \psi_i y^i$ are relatively prime polynomials in $\mathcal{F}^\dagger[y]$ and one of the coefficients ϕ_0, ψ_0 is 1. Since $(\phi\eta)^r \varepsilon \mathcal{G}$, for any automorphism τ of $\mathcal{G}^\dagger$ over $\mathcal{G}$ we may write $\tau(\phi\eta) = e\phi\eta$ where e is some r-th root of 1. As $\tau\theta = \theta$, we have $\sum \tau\phi_i \cdot e^i(\phi\eta)^i / \sum \tau\psi_i \cdot e^i(\phi\eta)^i = \sum \phi_i(\phi\eta)^i / \sum \psi_i(\phi\eta)^i$, so that $(\sum \tau\phi_i \cdot e^i(\phi\eta)^i)(\sum \psi_i(\phi\eta)^i) = (\sum \phi_i(\phi\eta)^i)(\sum \tau\psi_i \cdot e^i(\phi\eta)^i)$. Because of the relative primeness mentioned above and the fact that ϕ_0 or ψ_0 is 1, we conclude that $\sum \tau\phi_i \cdot e^i(\phi\eta)^i = \sum \phi_i(\phi\eta)^i$, $\sum \tau\psi_i \cdot e^i(\phi\eta)^i = \sum \psi_i(\phi\eta)^i$, so that $\tau\phi_i = e^{-i}\phi_i$, $\tau\psi_i = e^{-i}\psi_i$. Consider any value of i which is not divisible by r; writing $i = qr + r'$, where $0 < r' < r$, we find that $\tau(\phi_i(\phi\eta)^{r'}) = e^{-i}\phi_i e^{r'}(\phi\eta)^{r'} = \phi_i(\phi\eta)^{r'}$. Since τ is any automorphism of $\mathcal{G}^\dagger$ over $\mathcal{G}$ this implies that $\phi_i(\phi\eta)^{r'} \varepsilon \mathcal{G}$. Letting ϕ' be an element of $\mathcal{F}^\dagger$ such that $\phi'^{r'} = \phi_i\phi^{r'}$, we see that $(\phi'\eta)^{r'} \varepsilon \mathcal{G}$; because of the minimal nature of r and the relation $0 < r' < r$, we conclude that $\phi' = 0$, whence $\phi_i = 0$. Similarly $\psi_i = 0$ for all i not divisible by r. On the other hand, if i is divisible by r then $\tau\phi_i = e^{-i}\phi_i = \phi_i$, so that $\phi_i \varepsilon \mathcal{G} \cap \mathcal{F}^\dagger = \mathcal{F}$, and similarly, $\psi_i \varepsilon \mathcal{F}$. It follows that $\theta \varepsilon \mathcal{F}\langle(\phi\eta)^r\rangle = \mathcal{F}\langle\alpha\rangle$, so that $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$.

Finally we consider the third case. It is apparent that the point $(1 : \eta : a_i^{-1}\delta_i\eta)^{-n} \prod (1 : \tau_j\eta : \tau_j(a_i^{-1}\delta_i\eta))$ of $W(g_2, g_3)$ is invariant under σ. Since σ is any automorhpism of $\mathcal{G}^\dagger$ over $\mathcal{F}^\dagger$ other than the identity, this point is rational over $\mathcal{F}^\dagger$. Because $\mathcal{F}^\dagger$ is algebraically closed and $W(g_2, g_3)$ is a complete curve in the projective plane, this point can be written in the form P^n, where P is a point of $W(g_2, g_3)$ which is rational over $\mathcal{F}^\dagger$. For this P we have $(1 : \eta : a_i^{-1}\delta_i\eta)^n P^n = \prod (1 : \tau_j\eta : \tau_j(a_i^{-1}\delta_i\eta))$, which is clearly rational over $\mathcal{G}$. Now η is transcendental over $\mathcal{F}^\dagger$, so that $(1 : \eta : a_i^{-1}\delta_i\eta)$ is

a generic point of the curve $W(g_2, g_3)$ over $\mathcal{F}^\dagger$; since P is rational over $\mathcal{F}^\dagger$ $(1:\eta:a_i^{-1}\delta_i\eta)P$ is also a generic point of $W(g_2, g_3)$ over $\mathcal{F}^\dagger$, whence $(1:\eta:a_i^{-1}\delta_i\eta)^n P^n$ is, too. Since the field of constants of $\mathcal{G}^\dagger$ is clearly $\mathcal{C}$, which is contained in $\mathcal{F}^\dagger$, it follows from Lemma 2 (§ 6) that $(1:\eta:a_i^{-1}\delta_i\eta)^n P^n = (1:\alpha:\beta)$, where α is weierstrassian over $\mathcal{F}^\dagger$ with invariants g_2, g_3. But by the above, $\alpha, \beta \in \mathcal{G}$, so that α is weierstrassian over $\mathcal{G} \cap \mathcal{F}^\dagger = \mathcal{F}$. Also, α is transcendental over $\mathcal{F}$, so that $\mathcal{G}$ is algebraic over $\mathcal{F}\langle\alpha\rangle$. This completes the reduction in the third and final case.

10. The proof continued: case of genus 0. We assume now, in addition to the hypothesis of Theorem 7, that $\mathcal{F}$ is algebraically closed. Since $\mathcal{G}$ is of transcendence degree 1 over $\mathcal{F}$, we may regard $\mathcal{G}$ as a field of algebraic functions of one variable over $\mathcal{F}$; furthermore, every automorphism of the differential field $\mathcal{G}$ over $\mathcal{F}$ is obviously an automorphism of the algebraic function field $\mathcal{G}$. It is a well-known theorem that if the genus of such an algebraic function field is greater than 1 then the group of its automorphisms is finite (for a proof in the general (abstract) case see Iwasawa and Tamagawa [7]). It follows that the algebraic function field $\mathcal{G}$ has genus 0 to 1. In the present section we treat the case of genus 0 and show that in this case there exists an element α, which is either primitive or exponential over $\mathcal{F}$, such that $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$. In the next section we shall treat the case of genus 1.

Assuming then that $\mathcal{G}$ has genus 0, we see that $\mathcal{G}$ is a purely transcendental extension of $\mathcal{F}$ (see for example Chevalley [2], Chapter II, § 2), that is, that there exists a single element θ transcendental over $\mathcal{F}$ such that $\mathcal{G} = \mathcal{F}(\theta)$. Since $\delta_i\theta \in \mathcal{G}$ for each i, there exist polynomials $P_1, \cdots, P_m$, $Q \in \mathcal{F}[y]$ with $Q \neq 0$ such that

$$\delta_i\theta = Q(\theta)^{-1}P_i(\theta), \qquad 1 \leq i \leq m. \tag{13}$$

It is obvious that θ is not a constant, so that $P_i \neq 0$ for at least one value of i. If $c \in \mathcal{C}$ is not a zero of Q nor of any nonzero P_i, and if $k = \max(\deg P_1, \cdots, \deg P_m, \deg Q)$, then $\bar{Q}(y) = Q(y^{-1} + c)y^k$ and the nonzero expression $\bar{P}_i(y) = -P_i(y^{-1} + c)y^k$ are polynomials of degree k; but if we let $\bar{\theta} = (\theta - c)^{-1}$, so that $\mathcal{G} = \mathcal{F}(\bar{\theta})$, then $\delta_i\bar{\theta} = \bar{Q}(\bar{\theta})^{-1}\bar{P}_i(\bar{\theta})\bar{\theta}^2$ $(1 \leq i \leq m)$. Therefore we lose no generality in assuming that $\deg P_i = 2 + \deg Q$ for all values of i such that $P_i \neq 0$; we assume too, as we obviously may, that $P_1, \cdots, P_m, Q$ have no common divisor and that the leading coefficient in Q is 1. We denote the degree of Q by d.

Every automorphism of a simple transcendental extension is given by a fractional linear substitution. Therefore if σ is any automorphism of the

differential field $\mathcal{G}$ over $\mathcal{F}$ then there exist elements $a_{11}, a_{12}, a_{21}, a_{22}$ in $\mathcal{F}$ such that

$$\sigma\theta = (a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}), \qquad | \sigma | \neq 0,$$

where $| \sigma | = a_{11}a_{22} - a_{12}a_{21}$. Applying σ to each side of (13) we find

$$\begin{aligned}(a_{11}\delta_i\theta + \delta_i a_{11} \cdot \theta + \delta_i a_{12})(a_{21}\theta + a_{22})^{-1} \\ - (a_{11}\theta + a_{12})(a_{21}\theta + a_{22})^{-2}(a_{21}\delta_i\theta + \delta_i a_{21} \cdot \theta + \delta_i a_{22}) \\ = Q((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}))^{-1}P_i((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12})),\end{aligned}$$

which we rewrite (using (13)) in the form

$$\begin{aligned}(14)\quad & Q((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}))(a_{21}\theta + a_{22})^{d} \cdot (| \sigma | P_i(\theta) + A_i(\theta)Q(\theta)) \\ & = Q(\theta)P_i((a_{21}\theta + a_{22})^{-1}(a_{11}\theta + a_{12}))(a_{21}\theta + a_{22})^{d+2},\end{aligned}$$

where

$$\begin{aligned}(15)\qquad A_i(y) = {} & (a_{21}\delta_i a_{11} - a_{11}\delta_i a_{21})y^2 + (a_{22}\delta_i a_{11} - a_{11}\delta_i a_{22} \\ & + a_{21}\delta_i a_{12} - a_{12}\delta_i a_{21})y + a_{22}\delta_i a_{12} - a_{12}\delta_i a_{22}.\end{aligned}$$

It follows from (14) that for each i

$$Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^{d} \cdot | \sigma | P_i(y)$$

is divisible by $Q(y)$, whence the polynomial

$$Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^{d}$$

is so divisible. It follows that the fractional linear transformation

$$(16)\qquad x \to (a_{21}x + a_{22})^{-1}(a_{11}x + a_{12})$$

permutes the zeros of Q. Since this must happen for each of the infinitely many automorphisms σ of $\mathcal{G}$ over $\mathcal{F}$, and since a fractional linear transformation is uniquely determined by its values at three points, Q can have no more than two distinct zeros.

Supose Q has two distinct zeros ξ_1, ξ_2, so that $Q = (y - \xi_1)^{h_1}(y - \xi_2)^{h_2}$, where $h_1 + h_2 = d$; we suppose for the sake of definiteness that $h_1 \leq h_2$. For each σ in the group of all automorphisms of $\mathcal{G}$ over $\mathcal{F}$ the transformation (16) permutes ξ_1, ξ_2. The subgroup of all automorphisms σ for which (16) leaves ξ_1 and ξ_2 invariant is obviously of index 2, and is therefore infinite. For every σ of this infinite subgroup we have $(a_{21}\xi_j + a_{22})^{-1}(a_{11}\xi_j + a_{12}) = \xi_j$, that is $a_{21}\xi_j^2 + (a_{22} - a_{11})\xi_j - a_{12} = 0$ $(j = 1, 2)$, so that

$$(17)\qquad (a_{21} : a_{11} - a_{22} : -a_{12}) = (1 : \xi_1 + \xi_2 : \xi_1\xi_2).$$

Now because $Q((a_{21}y+a_{22})^{-1}(a_{11}y+a_{12}))(a_{21}y+a_{22})^d$ is divisible by $Q(y)$ and obviously has the same degree as $Q(y)$, the quotient of these two polynomials is in $\mathcal{F}$; since $Q(y)=(y-\xi_1)^{h_1}(y-\xi_2)^{h_2}$, an easy computation shows that this quotient is $(a_{11}-a_{21}\xi_1)^{h_1}(a_{11}-a_{21}\xi_2)^{h_2}$. But

$$(a_{11}-a_{21}\xi_1)(a_{11}-a_{21}\xi_2)=|\,\sigma\,|,$$

so that

$$Q((a_{21}y+a_{22})^{-1}(a_{11}y+a_{12}))(a_{21}y+a_{22})^d=|\,\sigma\,|^{h_1}(-a_{21}\xi_2+a_{11})^{h_1-h_2}Q(y).$$

It follows from (14) that

$$\begin{aligned}|\,\sigma\,|^{h_1}(-a_{21}\xi_2+a_{11})^{h_2-h_1}(|\,\sigma\,|\,P_i(y)+A_i(y)Q(y))\\ =P_i((a_{21}y+a_{22})^{-1}(a_{11}y+a_{12}))(a_{21}y+a_{22})^{d+2}.\end{aligned}$$

Replacing y by ξ_j we find that

$$|\,\sigma\,|^{h_1+1}(-a_{21}\xi_2+a_{11})^{h_2-h_1}P_i(\xi_j)=P_i(\xi_j)(a_{21}\xi_j+a_{22})^{d+2}.$$

Now, for at least one value of i, $P_i(\xi_j)\neq 0$, for otherwise $P_1,\cdots,P_m,Q$ would have the common factor $y-\xi_j$. Therefore

$$|\,\sigma\,|^{h_1+1}(-a_{21}\xi_2+a_{11})^{h_2-h_1}=(a_{21}\xi_j+a_{22})^{d+2}.$$

Since the left member here is the same for both values of j, the same must be true for the right member, that is $(a_{21}\xi_1+a_{22})^{d+2}=(a_{21}\xi_2+a_{22})^{d+2}$, so that $a_{21}\xi_1+a_{22}=\mu(a_{21}\xi_2+a_{22})$, where μ is one of the $(d+2)$-th roots of unity. But this equation and (17) together admit only a finite number of solutions $(a_{11}:a_{12}:a_{21}:a_{22})$. This contradicts the infinite number of possibilities for σ, and proves that Q can not have two distinct zeros.

Suppose now that Q has precisely one zero ξ. If we set $\bar{\theta}=\theta-\xi$, so that $\mathcal{G}=\mathcal{F}(\bar{\theta})$, then $\delta_i\bar{\theta}=\bar{\theta}^{-d}(P_i(\bar{\theta}+\xi)-\delta_i\xi\cdot\bar{\theta}^d)=\bar{\theta}^{-d}\bar{P}_i(\bar{\theta})$, where $\bar{P}_i(y)=P_i(y+\xi)-\delta_i\xi\cdot y^d$; whenever $P_i\neq 0$ then $\bar{P}_i\neq 0$ and is of degree $d+2$; if $P_i=0$ then, choosing some j such that P_j is not divisible by $y-\xi$ and letting P_{ji} denote the polynomial obtained by replacing each coefficient ϕ in P_j by $\delta_i\phi$, we find that

$$\begin{aligned}0&=\delta_j(Q(\theta)^{-1}P_i(\theta))=\delta_j\delta_i\theta=\delta_i\delta_j\theta\\&=\delta_i(Q(\theta)^{-1}P_j(\theta))=\delta_i((\theta-\xi)^{-d}P_j(\theta))\\&=-d(\theta-\xi)^{-d-1}((\theta-\xi)^{-d}P_i(\theta)-\delta_i\xi)P_j(\theta)\\&\quad+(\theta-\xi)^{-d}(P_{ji}(\theta)+P'_j(\theta)(\theta-\xi)^{-d}P_i(\theta))\\&=d(\theta-\xi)^{-d-1}\delta_i\xi\cdot P_j(\theta)+(\theta-\xi)^{-d}P_{ji}(\theta),\end{aligned}$$

so that $d\delta_i\xi \cdot P_j$ is divisible by $y - \xi$, whence $\delta_i\xi = 0$, and $\bar{R}_i = 0$. Therefore we lose no generality in assuming that $Q = y^d$. For every automorphism σ of $\mathcal{G}$ over $\mathcal{F}$ the transformation (16) then leaves 0 invariant, so that $a_{12} = 0$, and

$$Q((a_{21}y + a_{22})^{-1}(a_{11}y + a_{12}))(a_{21}y + a_{22})^d = a_{11}{}^d y^d = a_{11}{}^d Q(y).$$

It follows from (14) that

$$(18)\quad a_{11}{}^d(|\sigma| P_i(y) + A_i(y)y^d) = P_i((a_{21}y + a_{22}))^{-1}a_{11}y)(a_{21}y + a_{22})^{d+2}.$$

Replacing y by 0 here we find that

$$a_{11}{}^d \mid \sigma \mid P_i(0) = P_i(0)a_{22}{}^{d+2}.$$

But $|\sigma| = a_{11}a_{22} \neq 0$, and $P_i(0) \neq 0$ for some i; therefore $a_{11}{}^{d+1} = a_{22}{}^{d+1}$, whence $a_{22} = \mu a_{11}$, where now μ is one of the $(d+1)$-th roots of unity. By (15) we thus find that

$$A_i(y) = (a_{21}\delta_i a_{11} - a_{11}\delta_i a_{21})y^2 = - a_{11}{}^2\delta_i(a_{11}{}^{-1}a_{21}) \cdot y^2,$$

so that (18) becomes

$$\mu P_i(y) - \delta_i(a_{11}{}^{-1}a_{21}) \cdot y^{d+2} = P_i((a_{11}{}^{-1}a_{21}y + \mu)^{-1}y)(a_{11}{}^{-1}a_{21}y + \mu)^{d+2}.$$

Equating coefficients of y here we obtain

$$(d + 2)P_i(0)a_{11}{}^{-1}a_{21} = P'_i(0)(\mu - 1).$$

Since $P_i(0) \neq 0$ for some i, and since we already know that $a_{12} = 0$ and $a_{22} = \mu a_{11}$, this contradicts the fact that the number of automorphisms σ is infinite. Therefore Q does not have a zero, so that $Q = 1$, and each P_i which is different from 0 has degree 2.

Thus we may write

$$P_i = p_{i0} + p_{i1}y + p_{i2}y^2 \qquad (p_{ij} \in \mathcal{F},\ p_{i2} \neq 0 \text{ if } P_i \neq 0).$$

From this an easy computation shows that

$$\begin{aligned}\delta_j\delta_i\theta = \delta_j p_{i0} + p_{j0}p_{i1} + (\delta_j p_{i1} + 2p_{j0}p_{i2} + p_{j1}p_{i1})\theta \\ + (\delta_j p_{i2} + p_{j2}p_{i1} + 2p_{j1}p_{i2})\theta^2 + 2p_{j2}p_{i2}\theta^3.\end{aligned}$$

Since $\delta_j\delta_i = \delta_i\delta_j$ this implies that

$$(19)\qquad \begin{cases} \delta_j p_{i0} + p_{j0}p_{i1} = \delta_i p_{j0} + p_{i0}p_{j1}, \\ \delta_j p_{i1} + 2p_{j0}p_{i2} = \delta_i p_{j1} + 2p_{i0}p_{j2}, \\ \delta_j p_{i2} + p_{j1}p_{i2} = \delta_i p_{j2} + p_{i1}p_{j2}. \end{cases}$$

Let σ_0 be any automorphism of $\mathcal{G}$ over $\mathcal{F}$ such that $\sigma_0\theta \neq \theta$. It is easy to verify that the conditions

$$\delta_i(\tfrac{1}{2}p_{j1} + p_{j2}\theta) = \delta_j(\tfrac{1}{2}p_{i1} + p_{i2}\theta)$$

hold for all i and j. These conditions imply that the differential ideal $[\delta_1 z + (\frac{1}{2}p_{11} + p_{12}\theta)z, \cdots, \delta_m z + (\frac{1}{2}p_{m1} + p_{m2}\theta)z]$ in $\mathcal{G}\{z\}$ does not contain z and thus has a zero $\zeta_1 \neq 0$; it is not difficult to see that we may take ζ_1 so that the field of constants of $\mathcal{G}\langle\zeta_1\rangle$ is $\mathcal{C}$. Similarly, there exists a zero $\zeta_2 \neq 0$ of the differential ideal

$$[\delta_1 z + (\tfrac{1}{2}p_{11} + p_{12}\sigma_0\theta)z, \cdots, \delta_m z + (\tfrac{1}{2}p_{m1} + p_{m2}\sigma_0\theta)z]$$

in $\mathcal{G}\langle\zeta_1\rangle\{z\}$ such that the field of constants of $\mathcal{G}\langle\zeta_1, \zeta_2\rangle$ is $\mathcal{C}$. For any pair θ_1, θ_2 of operators of the form $\delta_1^{i_1} \cdots \delta_m^{i_m}$ we write

$$W_{\theta_1,\theta_2}(z_1, z_2) = \theta_1 z_1 \cdot \theta_2 z_2 - \theta_2 z_1 \cdot \theta_1 z_1.$$

It is easy to verify that

$$W_{1,\delta_i}(\zeta_1, \zeta_2) = p_{i2}\zeta_1\zeta_2(\theta - \sigma_0\theta), \tag{20}$$

which is different from 0 for at least one value of i, and that

$$W_{\theta_1,\theta_2}(\zeta_1, \zeta_2)\zeta_1^{-1}\zeta_2^{-1}(\theta - \sigma_0\theta)^{-1} \in \mathcal{F}$$

for every pair θ_1, θ_2 of operators of order ≤ 2. Therefore (Kolchin [6], § 3) $\mathcal{H} = \mathcal{F}\langle\zeta_1, \zeta_2\rangle$ is a Picard-Vessiot extension of $\mathcal{F}$. Of course $\mathcal{G} \subseteq \mathcal{H}$.

We denote the group of all automorphisms of $\mathcal{H}$ over $\mathcal{F}$ by $\mathfrak{H}$. By the Picard-Vessiot theory (Kolchin [6]) each element $\tau \in \mathfrak{H}$ may be identified with an element $(b_{ij}) = (b_{ij})_{1 \leq i \leq 2,\ 1 \leq j \leq 2}$ of an algebraic matric group over $\mathcal{C}$ by means of equations

$$\tau\zeta_j = b_{1j}\zeta_1 + b_{2j}\zeta_2 \qquad (j = 1, 2).$$

Let σ be any automorphism of $\mathcal{G}$ over $\mathcal{F}$ distinct from the identity and σ_0; σ can (Kolchin [6], § 3, Theorem 3) be extended to an element $\tau \in \mathfrak{H}$. Writing $\tau = (b_{ij})$ we have

$$\begin{aligned} 0 = \tau 0 &= \tau(\delta_i\zeta_1 + (\tfrac{1}{2}p_{i1} + p_{i2}\theta)\zeta_1) \\ &= b_{11}\delta_i\zeta_1 + b_{21}\delta_i\zeta_2 + (\tfrac{1}{2}p_{i1} + p_{i2}\sigma\theta)(b_{11}\zeta_1 + b_{21}\zeta_2) \\ &= -b_{11}(\tfrac{1}{2}p_{i1} + p_{i2}\theta)\zeta_1 - b_{21}(\tfrac{1}{2}p_{i1} + p_{i2}\sigma_0\theta)\zeta_2 \\ &\qquad + (\tfrac{1}{2}p_{i1} + p_{i2}\sigma\theta)(b_{11}\zeta_1 + b_{21}\zeta_2) \\ &= b_{11}p_{i2}(\sigma\theta - \theta)\zeta_1 + b_{21}p_{i2}(\sigma\theta - \sigma_0\theta)\zeta_2; \end{aligned}$$

since $\det(b_{ij}) \neq 0$ this implies that $b_{11}b_{21} \neq 0$ and $\zeta_1^{-1}\zeta_2 \in \mathcal{G}$.

A straightforward computation shows that $\delta_i(\zeta_1\zeta_2(\theta-\sigma_0\theta))=0$ $(1\leq i\leq m)$, so that $\zeta_1\zeta_2(\theta-\sigma_0\theta)=c\,\varepsilon\,\mathcal{C}$. From (20) we see at once that $W_{1,\delta_i}(\zeta_1,\zeta_2)=cp_{i2}\,\varepsilon\,\mathcal{F}$, so that

$$\det(b_{ij})\cdot W_{1,\delta_i}(\zeta_1,\zeta_2)=W_{1,\delta_i}(\tau\zeta_1,\tau\zeta_2)=\tau W_{1,\delta_i}(\zeta_1,\zeta_2)=W_{1,\delta_i}(\zeta_1,\zeta_2)$$

whence $\det(b_{ij})=1$ for all $\tau\,\varepsilon\,\mathfrak{H}$. Also, $\zeta_1^{-2}W_{1,\delta_i}(\zeta_1,\zeta_2)=\delta_i(\zeta_1^{-1}\zeta_2)\,\varepsilon\,\mathcal{G}$, so that $\zeta_1^{2}\,\varepsilon\,\mathcal{G}$, whence $\zeta_1^2,\zeta_1\zeta_2,\zeta_2^2\,\varepsilon\,\mathcal{G}$. Since $\delta_i\zeta_1^2=-2\zeta_1^2(\frac{1}{2}p_{i1}+p_{i2}\theta)$, it follows that $\theta\,\varepsilon\,\mathcal{F}\langle\zeta_1^2,\zeta_1\zeta_2,\zeta_2^2\rangle$, so that $\mathcal{G}=\mathcal{F}\langle\zeta_1^2,\zeta_1\zeta_2,\zeta_2^2\rangle$. Therefore $\partial^0\mathcal{H}/\mathcal{F}=\partial^0\mathcal{G}/\mathcal{F}=1$, whence $\dim\mathfrak{H}=1$.

It follows that the component of the identity $\mathfrak{H}^0$ is reducible either to diagonal form or else to special triangular form, that is, there exist two linear combinations ω_1, ω_2 of ζ_1, ζ_2 over $\mathcal{C}$, which are linearly independent over constants, such that either for every $\tau\,\varepsilon\,\mathfrak{H}^0$ there exists a nonzero $b\,\varepsilon\,\mathcal{C}$ for which $\tau\omega_1=b\omega_1$, $\tau\omega_2=b^{-1}\omega_2$, or else for every $\tau\,\varepsilon\,\mathfrak{H}^0$ there exists a $b\,\varepsilon\,\mathcal{C}$ for which $\tau\omega_1=\omega_1$, $\tau\omega_2=b\omega_1+\omega_2$.

In the former case $\tau(\omega_j^{-1}\delta_i\omega_j)=\omega_j^{-1}\delta_i\omega_j$ for every $\tau\,\varepsilon\,\mathfrak{H}^0$ so that $\omega_j^{-1}\delta_i\omega_j$ is algebraic over $\mathcal{F}$; but

$$\omega_j^{-1}\delta_i\omega_j=\tfrac{1}{2}\omega_j^{-2}\delta_i(\omega_j^2)\;\varepsilon\;\mathcal{F}\langle\zeta_1^2,\zeta_1\zeta_2,\zeta_2^2\rangle=\mathcal{G}$$

and $\mathcal{F}$ is relatively algebraically closed in $\mathcal{G}$, so that $\omega_j^{-1}\delta_i\omega_j\,\varepsilon\,\mathcal{F}$, that is, ω_j is exponential over $\mathcal{F}$. Therefore for every τ in $\mathfrak{H}$ there exists a nonzero $b\,\varepsilon\,\mathcal{C}$ such that $\tau\omega_1=b\omega_1$, $\tau\omega_2=b^{-1}\omega_2$. Consequently $\omega_1\omega_2$ is invariant under every $\tau\,\varepsilon\,\mathfrak{H}$ and belongs to $\mathcal{F}$, so that

$$\mathcal{G}=\mathcal{F}\langle\zeta_1^2,\zeta_1\zeta_2,\zeta_2^2\rangle=\mathcal{F}\langle\omega_1^2,\omega_1\omega_2,\omega_2^2\rangle=\mathcal{F}\langle\omega_1^2\rangle.$$

Setting $\alpha=\omega_1^2$ we see that $\alpha^{-1}\delta_i\alpha=2\omega_1^{-1}\delta_i\omega_1\,\varepsilon\,\mathcal{F}$, so that α is exponential over $\mathcal{F}$, and also that $\mathcal{G}=\mathcal{F}\langle\alpha\rangle$.

In the latter case $\tau\omega_1=\omega_1$ for every $\tau\,\varepsilon\,\mathfrak{H}^0$, so that ω_1 is algebraic over $\mathcal{F}$ whence, since $\omega_1^2\,\varepsilon\,\mathcal{F}\langle\zeta_1^2,\zeta_1\zeta_2,\zeta_2^2\rangle=\mathcal{G}$, we have $\omega_1^2\,\varepsilon\,\mathcal{F}$. Since $W_{1,\delta_i}(\zeta_1,\zeta_2)=cp_{i2}\,\varepsilon\,\mathcal{F}$, we also have $W_{1,\delta_i}(\omega_1,\omega_2)\,\varepsilon\,\mathcal{F}$, so that $\delta_i(\omega_1^{-1}\omega_2)=\omega_1^{-2}W_{1,\delta_i}(\omega_1,\omega_2)\,\varepsilon\,\mathcal{F}$. Therefore if we set $\alpha=\omega_1^{-1}\omega_2$ then α is primitive over $\mathcal{F}$ and $\mathcal{G}=\mathcal{F}\langle\omega_1^2,\omega_1\omega_2,\omega_2^2\rangle=\mathcal{F}\langle\omega_1^2,\omega_1^2\alpha,\omega_1^2\alpha^2\rangle=\mathcal{F}\langle\alpha\rangle$.

This completes the treatment of the case of genus 0.

11. The proof concluded. Case of genus 1. We consider now the remaining case in which $\mathcal{F}$ is algebraically closed and $\mathcal{G}$ is of genus 1. It is known (for example see Chevalley [2], Chapter II, § 3) that in this case there exist two elements α,β in $\mathcal{G}$ such that $\mathcal{G}=\mathcal{F}(\alpha,\beta)$ and $\beta^2=P(\alpha)$,

where P is a cubic polynomial in $\mathcal{F}[y]$ which does not have a multiple root. Replacing α, β by suitable elements $a\alpha + b$, $c\beta (a, b, c \in \mathcal{F})$, we lose no generality in supposing that

$$(21) \qquad \beta^2 = 4\alpha^3 - g_2\alpha - g_3,$$

where $g_2 \in \mathcal{C}$, $g_3 \in \mathcal{F}$, and $27g_3{}^2 - g_2{}^3 \neq 0$.

We shall prove that then $g_3 \in \mathcal{C}$, $\mathcal{G} = \mathcal{F}\langle\alpha\rangle$, and there exist elements $a_1, \cdots, a_m \in \mathcal{F}$ (not all 0) such that

$$(22) \qquad (\delta_i\alpha)^2 = a_i{}^2(4\alpha - g_2\alpha - g_3), \qquad (1 \leq i \leq m$$

(so that α is weierstrassian over $\mathcal{F}$). This will complete the proof of Theorem 7.

We begin by observing that α is transcendental over $\mathcal{F}$; since the field of constants of $\mathcal{G}$ is $\mathcal{C}$, α is not a constant. If $\delta_i\alpha = 0$ then (22) holds with $a_i = 0$. Let i be any index such that $\delta_i\alpha \neq 0$; in what follows we shall keep i fixed, and for every element ξ of $\mathcal{G}$ we shall denote $\delta_i\xi$ by ξ'.

Clearly there exist polynomials $A, B, C \in \mathcal{F}[y]$, without common divisor and with the leading coefficient in C equal to 1, such that

$$(23) \qquad \alpha' = \frac{A(\alpha) + B(\alpha)\beta}{C(\alpha)}.$$

Applying δ_i to both members of (21) we find that $2\beta\beta' = (12\alpha^2 - g_2)\alpha' - g'_3$; from this, (21), and (23) we obtain

$$(24) \quad \beta' = \frac{(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) + ((12\alpha^2 - g_2)A(\alpha) - g'_3C(\alpha))\beta}{2C(\alpha)(4\alpha^3 - g_2\alpha - g_3)}.$$

Now $(1 : \alpha : \beta)$ is a point of the group variety $W(g_2, g_3)$ defined in § 6. Since there exist infinitely many automorphisms over $\mathcal{F}$ of the differential field $\mathcal{G}$, there exist infinitely many such automorphisms σ such that

$$(25) \qquad (1 : \sigma\alpha : \sigma\beta) = (1 : a : b)(1 : \alpha : \beta),$$

where $a, b \in \mathcal{F}$ and $(1 : a : b) \in W(g_2, g_3)$,[10] that is, such that

---

[10] This follows from the known fact that, in the group of all automorphisms of the algebraic function field G, the subgroup of those automorphisms for which an equation of the form (25) holds is of finite index. To see this observe that this subgroup acts transitively on the set of all places of the function field G, and that there exists a place $\mathfrak{p}$ at which α has a pole of order 2 and β has a pole of order 3; the above fact is then a consequence of a second fact, namely that the group of all automorphisms of the function field G which leave $\mathfrak{p}$ invariant is finite. This second fact is in turn an easy

$$(26)\quad \begin{cases} \sigma\alpha = -\alpha - a + \frac{1}{4}\left(\dfrac{\beta - b}{\alpha - a}\right)^2, \\ \sigma\beta = -\frac{1}{2}(\beta + b) + \frac{3}{2}\,\dfrac{\beta - b}{\alpha - a} - \frac{1}{4}\left(\dfrac{\beta - b}{\alpha - a}\right)^3. \end{cases}$$

These equations may, with the help of (21), be rewritten in the form

$$(27)\quad \begin{cases} \sigma\alpha = \dfrac{4a\alpha^2 + (4a^2 - g_2)\alpha - (g_2 a + 2g_3) - 2b\beta}{4(\alpha - a)^2}, \\ \sigma\beta = \dfrac{(4\alpha^3 + 12a\alpha^2 - 3g_2\alpha - g_2 a - 4g_3)b + ((-12a^2 + g_2)\alpha - 4a^3 + 3g_2 a + 4g_3)\beta}{4(\alpha - a)^3}. \end{cases}$$

Applying δ_i to the first equation (26), and making use of (21), (23), and (24), we obtain

$$(\sigma\alpha)' = (U + V\beta)4^{-1}(\alpha - a)^{-1}(4\alpha^3 - g_2\alpha - g_3)^{-1}C(\alpha)^{-1},$$

where

$$\begin{aligned} (28)\quad U = {} & -4a'(\alpha - a)^3(4\alpha^3 - g_2\alpha - g_3)C(\alpha) - 4(\alpha - a)^3(4\alpha^3 - g_2\alpha - g_3)A(\alpha) \\ & + (\alpha - a)(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)A(\alpha) - g'_3(\alpha - a)(4\alpha^3 - g_2\alpha - g_3)C(\alpha) \\ & - b(\alpha - a)(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ & + ((12a^2 - g_2)a' - g'_3)(\alpha - a)(4\alpha^3 - g_2\alpha - g_3)C(\alpha) \\ & + 2b(4\alpha^3 - g_2\alpha - g_3)(A(\alpha) - a'C(\alpha)) - 2(4\alpha^3 - g_2\alpha - g_3)^2B(\alpha), \end{aligned}$$

and

$$\begin{aligned} (29)\quad V = {} & -4(\alpha - a)^3(4\alpha^3 - g_2\alpha - g_3)B(\alpha) + (\alpha - a)(12\alpha^2 - g_2)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ & - 2b'(\alpha - a)(4\alpha^3 - g_2\alpha - g_3)C(\alpha) - b(\alpha - a)((12\alpha^2 - g_2)A(\alpha) - g'_3C(\alpha)) \\ & - 2(4\alpha^3 - g_2\alpha - g_3)(A(\alpha) - a'C(\alpha)) + 2b(4\alpha^3 - g_2\alpha - g_3)B(\alpha). \end{aligned}$$

On the other hand, by (23)

$$\sigma(\alpha') = \frac{A(\sigma\alpha) + B(\sigma\alpha)\sigma\beta}{C(\sigma\alpha)}.$$

Since $(\sigma\alpha)' = \sigma(\alpha')$ we therefore find, with the help of (27), that

---

consequence of the Riemann-Roch theorem; indeed if σ_0 is any automorphism which leaves $\mathfrak{p}$ invariant then $\sigma_0\alpha$ has a pole of order 2 at $\mathfrak{p}$ and $\sigma_0\beta$ has a pole of order 3 there, whence (for example see Chevalley [2], chapter II, corollary to theorem 6) $\sigma_0\alpha = c_1\alpha + c_2$, $\sigma_0\beta = c_3\beta + c_4\alpha + c_5$, where $c_1, \cdots, c_5 \in \mathcal{F}$ and $c_1c_3 \neq 0$; since σ_0 must preserve equation (21), an easy computation shows that $c_2 = c_4 = c_5 = 0$, $c_1^3 - c_3^2 = 0$, $g_2(c_1 - c_3^2) = 0$, $g_3(c_3^2 - 1) = 0$, so that there are only a finite number of possibilities for σ_0. For this short proof I am indebted to M. Rosenlicht.

$$(30)\qquad 4(\alpha-a)^3(4\alpha^3-g_2\alpha-g_3)C(\alpha)\Big[A(W-\frac{b\beta}{2(\alpha-a)^2}) \\ +B(W-\frac{b\beta}{2(\alpha-a)^2}) \\ \times\frac{(4\alpha^3+12a\alpha^2-3g_2\alpha-g_2a-4g_3)b+((-12a^2+g_2)\alpha-4a^3+3g_2a+4g_3)\beta}{4(\alpha-a)^3}\Big] \\ =C(W-\frac{b\beta}{2(\alpha-a)^2})(U+V\beta),$$

where

$$W=\frac{4a\alpha^2+(4a^2-g_2)\alpha-(g_2a+2g_3)}{4(\alpha-a)^2}.$$

Because of (21) the left hand member here is a linear combination of 1 and β with coefficients which belong to $\mathcal{F}(\alpha)$, have denominators that are powers of $\alpha-a$, and have numerators that are divisible by $(4\alpha^3-g_2\alpha-g_3)C(\alpha)$. The right hand member can also be expressed as a linear combination of 1 and β, the coefficient of 1 being

$$(31)\qquad \sum_{j\geq 0}\frac{1}{(2j)!}C^{(2j)}(W)\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j}{2^{2j}(\alpha-a)^{4j}}U \\ -\sum_{j\geq 0}\frac{1}{(2j+1)!}C^{(2j+1)}(W)\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^{j+1}}{2^{2j+1}(\alpha-a)^{4j+2}}bV,$$

and the coefficient of β being

$$(32)\qquad \sum_{j\geq 0}\frac{1}{(2j)!}C^{(2j)}(W)\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j}{2^{2j}(\alpha-a)^{4j}}V \\ -\sum_{j\geq 0}\frac{1}{(2j+1)!}C^{(2j+1)}(W)\frac{(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j}{2^{2j+1}(\alpha-a)^{4j+2}}bU.$$

Therefore (31) and (32) are both expressible as quotients in which the denominator is a power of $\alpha-a$ and the numerator is a polynomial in $\mathcal{F}[\alpha]$ divisible by $(4\alpha^3-g_2\alpha-g_3)C(\alpha)$.

Observing from (28) that U is divisible by $4\alpha^3-g_2\alpha-g_3$, and from (29) that each term of V is so divisible except for

$$-b(\alpha-a)((12\alpha^2-g_2)A(\alpha)-g'_3C(\alpha)),$$

and recalling that α is transcendental over $\mathcal{F}$, we conclude, on substituting for α in (32) any root e of $4y^3-g_2y-g_3$, that

$$C\left(\frac{4ae^2+(4a^2-g_2)e-(g_2a+2g_3)}{4(e-a)^2}\right) \\ \times b(e-a)((12e^2-g_2)A(e)-g'_3C(e))=0.$$

Since this is true for infinitely many points $(1:a:b)$ of the curve $W(g_2, g_3)$, this implies that $(12e^2 - g_2)A(e) - g'_3C(e) = 0$. Because this equation holds for each of the three roots e of $4y^3 - g_2y - g_3$ we conclude that

$$(33) \qquad (12y^2 - g_2)A(y) - g'_3C(y) \equiv 0 \qquad (\mathrm{mod}\ 4y^3 - g_2y - g_3).$$

Returning now to (31) we see that if r denotes the degree of $C(y)$ and if we multiply (31) by $(\alpha - a)^{2r}$ then we obtain a polynomial in $\mathfrak{F}[\alpha]$ divisible by $C(\alpha)$, that is, we have a congruence in $\mathfrak{F}[\alpha]$ of the form

$$LU + MbV \equiv 0 \qquad (\mathrm{mod}\ C(\alpha)).$$

Since every subgroup of $W(g_2, g_3)$ which contains $(1:a:b)$ also contains $1:a:-b) = (1:a:b)^{-1}$, this congruence continues to hold if we replace b by $-b$. Using the two congruences (one with b and one with $-b$), and observing from (28) and (29) that

$$\begin{aligned} U \equiv {} & (-4(\alpha - a)^3 + (\alpha - a)(12\alpha^2 - g_2) + 2b)(4\alpha^3 - 2g_2\alpha - g_3)A(\alpha) \\ & - (b(\alpha - a)(12\alpha^2 - g_2) + 2(4\alpha^3 - g_2\alpha - g_3))(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ & \qquad (\mathrm{mod}\ C(\alpha)) \end{aligned}$$

and

$$\begin{aligned} V \equiv {} & -(b(\alpha - a)(12\alpha^2 - g_2) + 2(4\alpha^3 - g_2\alpha - g_3))A(\alpha) \\ & + (-4(\alpha - a)^3 + (\alpha - a)(12\alpha^2 - g_2) + 2b)(4\alpha^3 - g_2\alpha - g_3)B(\alpha) \\ & \qquad (\mathrm{mod}\ C(\alpha)), \end{aligned}$$

we obtain the following two congruences in $\mathfrak{F}[\alpha]$, in which b no longer appears:

$$\begin{aligned} (34) \quad & [\sum_{j \geqq 0} \frac{1}{(2j)!2^{2j}} C^{(2j)}(W)(\alpha - a)^{2r-4j} \\ & \quad \times (4a^3 - g_2a - g_3)^j(4\alpha^3 - g_2\alpha - g_3)^j(12\alpha^2 - g_2 - 4(\alpha - a)^2) \\ & + \sum_{j \geqq 0} \frac{1}{(2j+1)!2^{2j+1}} C^{(2j+1)}(W)(\alpha - a)^{2r-4j-2} \\ & \quad \times (4a^3 - g_2a - g_3)^{j+1}(4\alpha^3 - g_2\alpha - g_3)^j(12\alpha^2 - g_2)](\alpha - a)A(\alpha) \\ & - [\sum_{j \geqq 0} \frac{1}{(2j)!2^{2j}} C^{(2j)}(W)(\alpha - a)^{2r-4j} \\ & \quad \times (4a^3 - g_2a - g_3)^j(4\alpha^3 - g_2\alpha - g_3)^{j+1} \\ & + \sum_{j \geqq 0} \frac{1}{(2j+1)!2^{2j+1}} C^{(2j+1)}(W)(\alpha - a)^{2r-4j-2} \\ & \quad \times (4a^3 - g_2a - g_3)^{j+1}(4\alpha^3 - g_2\alpha - g_3)^{j+1}]2B(\alpha) \\ & \qquad \equiv 0\ (\mathrm{mod}\ C(\alpha)), \end{aligned}$$

and

$$(35)\quad [\sum_{j\geqq 0}\frac{1}{(2j)\,!2^{2j}}C^{(2j)}(W)\,(\alpha-a)^{2r-4j} \times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j$$
$$+\sum_{j\geqq 0}\frac{1}{(2j+1)\,!2^{2j+1}}C^{(2j+1)}(W)\,(\alpha-a)^{2r-4j-2} \times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^{j+1}]2A(\alpha)$$
$$-[\sum_{j\geqq 0}\frac{1}{(2j)\,!2^{2j}}C^{(2j)}(W)\,(\alpha-a)^{2r-4j} \times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^j(12\alpha^2-g_2)$$
$$+\sum_{j\geqq 0}\frac{1}{(2j+1)\,!2^{2j+1}}C^{(2j+1)}(W)\,(\alpha-a)^{2r-4j-2}$$
$$\times(4a^3-g_2a-g_3)^j(4\alpha^3-g_2\alpha-g_3)^{j+1}(12\alpha^2-g_2-4(\alpha-a)^2)](\alpha-a)B(\alpha) \equiv 0\ (\mathrm{mod}\ C(\alpha)).$$

Since the congruences (34) and (35) hold for infinitely many points $(1:a:b)\ \varepsilon\ \mathbf{W}(g_2,g_3)$ with $a,b\ \varepsilon\ \mathfrak{F}$, that is for infinitely many elements $a\ \varepsilon\ \mathfrak{F}$, they must hold for all elements a. In particular, they hold for a equal to α. Now the leading coefficient in $C(y)$ is 1; therefore, when we replace a by α, $C^{(k)}(W)(\alpha-a)^{2r-2k}$ becomes $(r\,!(r-k)\,!)2^{r-k}(4\alpha^3-g_2\alpha-g_3)^{r-k}$. Consequently, when a is replaced by α, (34) becomes

$$-[\sum_{j\geqq 0}\binom{r}{2j}2^{r-4j}(4\alpha^2-g_2\alpha-g_3)^{r+1} +\sum_{j\geqq 0}\binom{r}{2j+1}2^{r-4j-2}(4\alpha^3-g_2\alpha-g_3)^{r+1}]2B(\alpha)\equiv 0\quad(\mathrm{mod}\ C(\alpha)),$$

so that

$$B(\alpha)\,(4\alpha^3-g_2\alpha-g_3)^{r+1}\equiv 0\qquad(\mathrm{mod}\ C(\alpha)).$$

In the same way, on replacing a by α in (35), we find that

$$A(\alpha)\,(4\alpha^3-g_2\alpha-g_3)^{r}\equiv 0\qquad(\mathrm{mod}\ C(\alpha)).$$

Since A, B, C have no common divisor, it follows that

$$(4y^3-g_2y-g_3)^{r+1}\equiv 0\qquad(\mathrm{mod}\ C(y)).$$

Therefore if C were of degree >0 then C would have a root e in common with $4y^3-g_2y-g_3$, and we would obtain, on replacing α by e in (35)

$$C\left(\frac{4ae^2+4(a^2-g_2)e-(g_2a+2g_3)}{4(e-a)^2}\right)(e-a)^{2r}$$
$$\times\ (2A(e)-(e-a)(12e^2-g_2)B(e))=0,$$

so that we would have $A(e)=0$, $B(e)=0$, contradicting the fact that A, B, C are without common divisor. Consequently

$$C(y)=1. \tag{36}$$

Now A and B are not both 0, for otherwise by (23) we would have $\delta_i\alpha=\alpha'=0$, contrary to assumption. Let p denote the maximum of the degrees of A and B, and denote the coefficients of y^p in A and B by c_A and c_B respectively. Suppose p were >0. Multiplying both sides of (30) by $(\alpha-a)^{2p}$, then expressing each side as a linear combination, over $\mathcal{F}(\alpha)$, of 1 and β, and then equating coefficients of 1, we would obtain an equation in $\mathcal{F}[\alpha]$; if in this equation we equated the coefficients, right and left of α^{3p+6} we would obtain

$$32c_A-(48b+32)c_B=0;$$

since this would hold for infinitely many points $(1:a:b)\ \varepsilon\ W(g_2,g_3)$, that is, for infinitely many values of b, we would have $c_A=c_B=0$, which is impossible. Therefore $p=0$, so that $A=c_A\ \varepsilon\ \mathcal{F}$, $B=c_B\ \varepsilon\ \mathcal{F}$. By (36) and (33) we further conclude that $c_A=0$ and $g'_3=0$, whence from (23) $\alpha'=c_B\beta$. Writing $c_B=a_i$ we see, by (21), that (22) holds whence $\mathcal{G}=\mathcal{F}(\alpha,\beta)=\mathcal{F}\langle\alpha\rangle$. To complete the proof of the theorem it remains to show that $g_3\ \varepsilon\ \mathcal{C}$. We have just seen that $\delta_ig_3=0$ for those values of i for which $\delta_i\alpha\neq0$; we must prove that $\delta_kg_3=0$ for all values of k such that $\delta_k\alpha=0$. For such i and k we have

$$\delta_k((\delta_i\alpha)^2)=2\delta_i\cdot\alpha\cdot\delta_k\delta_i\alpha=2\delta_i\alpha\cdot\delta_i\delta_k\alpha=0$$

so that, because of (22),

$$0=\delta_k(a_i^2(4\alpha^3-g_2\alpha-g_3))=2a_i\cdot\delta_ka_i\cdot(4\alpha^3-g_2\alpha-g_3)+a_i^2(-\delta_kg_3);$$

since α is transcendental over $\mathcal{F}$, this implies that $\delta_ka_i=0$ and $\delta_kg_3=0$. Thus $g_3\ \varepsilon\ \mathcal{C}$, and the proof of Theorem 7 is complete.

COLUMBIA UNIVERSITY.

REFERENCES.

[1] N. Bourbaki, "*Algèbre*," Chapters IV-V (Actualités scientifiques et industrielles 1102), Hermann et Cie., Paris, 1950.

[2] C. Chevalley, "*Introduction to the theory of algebraic functions of one variable*," (Mathematical Surveys VI), American Mathematical Society, New York, 1951.

[3] E. R. Kolchin, "Algebraic matric groups and the Picard-Vessiot theory of homogeneous linear ordinary differential equations," *Annals of Mathematics*, vol. 49 (1948), pp. 1-42.

[4] ———, "Existence theorems connected with the Picard-Vessiot theory of homogeneous linear ordinary differential equations," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 927-932.

[5] ———, "On certain concepts in the theory of algebraic matric groups," *Annals of Mathematics*, vol. 49 (1948), pp. 774-789.

[6] ———, "Picard-Vessiot theory of partial differential fields," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 596-603.

[7] K. Iwasawa and T. Tamagawa, "On the group of automorphisms of a function field," *Journal of the Mathematical Society of Japan*, vol. 3 (1951), pp. 137-147.

[8] J. F. Ritt, "*Differential algebra*," (American Mathematical Society Colloquium Publications, vol. 33), New York, 1950.

[9] A. Weil, "*Foundations of algebraic geometry*," (ibid., vol. 29), New York, 1946.

[10] ———, "*Variétés abeliennes et courbes algébriques* (Actualités scientifiques et industrielles 1064), Hermann et Cie., Paris, 1948.

NETWORKS SATISFYING MINIMALITY CONDITIONS.*[1,2]

By R. DUNCAN LUCE.

1. **Introduction.** A *network* N is a system of two finite sets M and $P \subset M \times M$, in which the elements $a, b, \cdots \varepsilon M$ are called the *nodes* and the elements $(ab), (ca), \cdots \varepsilon P$ are called the *links* of N. The number of nodes is denoted by m and the number of links by $p(N)$. If (ab) is a link of N, a is called the *initial node* and b the *end node* of the link (ab). Thus, a network is a binary relation over a finite set and is also a finite oriented graph in which there is at most one oriented arc from one node to another. Our viewpoint is primarily that of graph theory rather than algebra.

Let I be the set of all links of the form $(aa), a \varepsilon M$. If $P \cap I = 0$, N is called *non-reflexive.* We shall, without further mention, take the word network to mean non-reflexive network.

A *subnetwork* N' of a network N, denoted $N' \subset N$, is any network with $M' \subset M$ and $P' \subset P$. If $M' = M$, we say N' is a complete subnetwork of N. If $N' \subset N$, $N - N'$ is the network with nodes M and links $P - P'$ and it is said to be formed from N by the removal of the links P'. If N' has but one link (ab) we write $N - N' = N - (ab)$. Similarly, the network with nodes $M - M'$ and links $P \cap [(M - M') \times (M - M')]$ is said to be formed from N by the removal of the nodes M' (and the incident links). N' is a *supernetwork* of N if N is a complete subnetwork of N'. We shall write in this case $N' = N + (N' - N)$ and say that N' is formed from N by adding the links $P' - P$ to N. If $N' - N$ contains but one link (ab), we write $N' = N + (ab)$.

A pair of links (ab) and (ba) is called an *arc ab,* and any network composed entirely of arcs is isomorphic to a graph and so is called a graph.

A *q-chain* from a to b, denoted (ab, q), is an ordered sequence of q links $(ac_1), (c_1c_2), \cdots, (c_qb)$ in which no node is repeated, except possibly $a = b$. In the latter case the chain is called an (oriented) *circuit.* A network is

* Received April 6, 1953.

[1] This work has been supported in part by the Signal Corps, the Air Materiel Command, and the Office of Naval Research, U. S. A.

[2] The author is indebted to Josiah Macy, Jr., for many helpful discussions about this work and for a critical reading of the manuscript.

connected if there is a chain from every node to every other node; otherwise it is *disconnected.* Maximal connected subnetworks are called *components.*

In a previous paper [4], which will be referred to as (A), the following definitions were introduced:

A network has *degree* 0 if it is not connected; it has *degree* k, $k > 0$, if there exists $N' \subset N$ such that $p(N') = k$ and $N - N'$ is disconnected, but $N - N''$ is connected for all $N'' \subset N$ such that $p(N'') < k$.

A network N is *k-minimal* if the degree of $N - (ab)$ is $k - 1$ for every $(ab) \varepsilon N$. If N is 1-minimal and connected, it is called *minimal.*

In this paper we are concerned with three independent results which are each related to k-minimality. The definition is extended in a natural way to disconnected networks in Section 2 and these networks are completely characterized by Theorem 1. It is worth mentioning that the characterization problem for connected networks appears to be far more difficult. (The principal result of (A) is the solution to that problem for $k = 1$). In Section 3, the principal result is Theorem 4 which states that in a network of degree k, there is a set of at least k chains from any node to any other node, no two of which have a common link. This result is a generalization of a close analogue to the well known theorem of Menger that between any two nodes of a graph without a cut-node there are at least two chains that have no intermediate nodes in common. In the final section we turn to a generalization of transitivity. Connectedness and transitivity are each such strong requirements that combined they single out but one network—the case $P = M \times M$—so, in the presence of connectedness, transitivity must be weakened to be of interest. We require that every chain exceeding h links is "short-circuited" by a link, and that no chain of h or fewer links is short-circuited. It is shown that these connected networks fall into three classes: one having but one member which is of degree 2, the set of minimal networks, and a set non-minimal networks of degree 1 whose connected subnetworks also have degree 1.

2. ($-k$)-minimal networks.[3] To extend the above definitions of degree and minimality to disconnected networks, we simply interchange the roles of connected and disconnected as follows:

A network N has *degree* $(-k)$, $k \geqq 0$, if there exists a connected supernetwork N' of N such that $p(N' - N) = k + 1$, but every supernetwork N'' such that $p(N'' - N) < k + 1$ is disconnected.

[3] The author is indebted to Anatol Holt who suggested this problem to him.

A network N is (-0)-*minimal* if N is disconnected and for every $(ab) \notin N, a, b \in N$, $N + (ab)$ is connected; it is $(-k)$-minimal, $k \geqq 1$, if for every $(ab) \notin N, a, b \in N$, $N + (ab)$ has degree $(-k+1)$.

Following Dirac's terminology ([1], p. 347), we shall call a network with every possible link present a *complete graph*, and any component which is a complete graph is simply called complete. A node which is neither an end node nor an initial node of any link is called an *isolated node.*

LEMMA 1. *If N is a $(-k)$-minimal network with $m \geqq 3$ and $k \geqq 2$ and a is an isolated node of N, then $N' = N - a$ is either $(-k+1)$-minimal or a complete graph.*

Proof. If N' is not a complete graph, then since $m \geqq 3$, there exist $b, c \in N'$ such that $(bc) \notin N'$. For any such b and c, consider $N^* = N' + (bc)$. Let $-q$ be the degree of N^*, then the lemma is proved if we show $q = k - 2$. Let U be any set of k links which connects $N + (bc)$, and observe, since a is isolated, there exist $e, f \in N'$ such that $(ea), (af) \in U$. Now, $U + (ef) - (ea) - (af)$ connects N^*, so $q \leqq k - 2$. Suppose $q < k - 2$ and let U' be a set of $q + 1$ links which connects N^*. U' is non-empty, for otherwise N^* is connected, whence $N + (ba)$ is connected by adding (ac), and this implies N *is* (-1)-minimal, which contradicts $k \geqq 2$. Let $(e'f') \in U'$, then $U' - (e'f') + (e'a) + (af')$ connects $N + (bc)$ using only $(q+1) + 1 < k$ links, which is a contradiction.

THEOREM 1. *A network N is $(-k)$-minimal if and only if either*

(i) *N is a graph which consists of $k+1$ complete components having no link between any pair, or*

(ii) *N consists of a set X of nodes which form $k+1$ complete components having no link between any pair and a complete component Y such that either*

1. $(xy) \in N$ *and* $(yx) \notin N$ *for all* $x \in X$ *and* $y \in Y$,

or

2. $(yx) \in N$ *and* $(xy) \notin N$ *for all* $x \in X$ *and* $y \in Y$.

Proof. The sufficiency is obvious.

The condition is clearly necessary for $k = 0$, so we restrict the proof to $k \geqq 1$. Let N' be any component of N. If N' is an isolated node, it is complete. If N' has more than one node, we show it is complete: If there

exist $a, b \varepsilon N'$ such that $(ab) \notin N'$, and if U is any set of k links which connects $N + (ab)$, then for any $(cd) \varepsilon U$, $U - (cd)$ connects $N + (cd)$, since N' is already connected. This contradicts the assumption that N is $(-k)$-minimal.

If N', N'' are two components of N, we show that if $a \varepsilon N'$, $b \varepsilon N''$, and $(ab) \varepsilon N$, then $(a'b') \varepsilon N$ for any $a' \varepsilon N'$, $b' \varepsilon N''$: Suppose $(a'b') \notin N$, and let U be any set of k links connecting $N + (a'b')$. U connects N since N' and N'' are complete and $(ab) \varepsilon N$, which contradicts the assumption that N is $(-k)$-minimal.

Since the components of N are complete and since if there is one link from N' to N'' there are all possible links, it is sufficient to prove the theorem for networks having no components with more than one node.

First, $m \geqq k + 1$, for if not N can be connected as a circuit on all nodes with fewer than $k + 1$ links. If $m = k + 1$, no links are present, for if there were N could again be connected as a circuit using no more than $m - 1 = k$ links. In this case, N satisfies part (i) of the statement.

For networks with $m \geqq k + 2$, an induction on m will be used to show part (ii) holds. For $m = 3$, it is clear this is the case. Suppose $m > 3$ and part (ii) holds for $m' = m - 1$. If N has an isolated node a, then by Lemma 1, $N - a$ is $(-k + 1)$-minimal, so by the induction hypothesis (ii) holds for $N - a$, since (i) cannot. Thus, there exists a node $d \varepsilon N - a$ such that for any other node $c \varepsilon N - a$, exactly one of (cd) and $(dc) \varepsilon N$. Suppose, without loss of generality, $(cd) \varepsilon N$. Then, $N + (ad)$ has degree $(-k)$ and $N + (da)$ has degree $(-k + 1)$, which is a contradiction, so N has no isolated nodes.

Divide the nodes of N into three classes: $X =$ set of initial nodes, $Y =$ set of end nodes, and $Z =$ set of nodes which are both initial and end nodes. Let these sets have q, p, and $m - q - p$ members respectively. It is simple to see that if $q = 0$ or $p = 0$ there is a connected subnetwork of N, which is impossible. Suppose $q \geqq p$.

Since the nodes of X terminate no links, at least q links will have to be added to N to produce a connected supernetwork. We shall now show that q links suffice. There are maximal subsets X_1 and Y_1 such that there is a $1:1$ correspondence $x_i \varepsilon X_1$, $y_i \varepsilon Y_1$, $i = 1, 2, \cdots, s$, and $(x_i y_i) \varepsilon N$. This follows from the fact that neither X nor Y are empty and from any $x \varepsilon X$ there is either a link to a $y \varepsilon Y$ or a chain via Z to a $y \varepsilon Y$. But in the latter case, $(xy) \varepsilon N$ for if not then N can be connected by the same set of links which connect $N + (xy)$.

The addition of the s links $(y_1 x_2)$, $(y_2 x_3)$, $\cdots$, $(x_s y_1)$ to N creates a

connected network on the nodes $X_1 + Y_1$. Let $\xi \varepsilon X - X_1$ and $\eta \varepsilon Y - Y_1$; then $(\xi\eta) \not\varepsilon N$, since X_1, Y_1 are maximal. Thus, if $\xi \varepsilon X - X_1$, there exists $y \varepsilon Y_1$ such that $(\xi y) \varepsilon N$, and if $\eta \varepsilon Y - Y_1$, there exists $x \varepsilon X_1$ such that $(x\eta) \varepsilon N$. Now, from each of the $p - s$ nodes of $Y - Y_1$ introduce links to the nodes of $X - X_1$ such that no two terminate on the same node; this is possible since $q \geqq p$. To each of the remaining nodes of $X - X_1$, if any, introduce a link from a node of Y_1. It is easy to see the resulting network is connected and that $s + (q - s) = q$ links have been added.

If $z \varepsilon Z$ and $y \varepsilon Y$, then $N + (yz)$ still requires the addition of q links to connect, so $Z = 0$ and $p = m - q$. If p were > 1, then for y_1, $y_2 \varepsilon Y$, $N + (y_1 y_2)$ would also require the addition of q links to connect; hence $p = 1$ and $q = k + 1$. Thus, $m = q + 1 = k + 2$.

If $p \geqq q$, a similar argument applies.

3. Analogue to Menger's theorem. In graph theory, a node of a connected graph is called a *cut-node* if its removal, along with the incident arcs, results in a graph having two or more components. We generalize this notion: a set of nodes of a connected network is called a *cut-set* if it is one of the smallest sets of nodes whose removal, along with the incident links, results in a disconnected network. If the cut-sets of a network each have κ members, we say the network has *index* κ. It is clear that every connected network has a unique index κ, that $1 \leqq \kappa \leqq m - 1$, and that a connected graph has a cut-node if and only if the index is 1.

The notions of index and degree are parallel with respect to the removal of nodes and links, and so presumedly their values cannot be completely independent. Our first result establishes a relation between them.

THEOREM 2. *Let a connected network on m nodes have degree k and index κ, then $\kappa \leqq k \leqq (m - 1 + \kappa)/2$.*

Proof. To show the left side of the inequality we prove: If a connected network N has degree $k < m - 1$ and $(a_j b_j)$, $j = 1, 2, \cdots, k$, are a set of links whose removal disconnects N, then there exists a set of nodes c_j, $j = 1, 2, \cdots, k$, with $c_j = a_j$ or b_j, such that their removal results in a disconnected network. The c_j are not necessarily distinct.

For $m = 2$ this is obvious.

Consider $m \geqq 3$ and $k = 1$. Let c and d be two nodes such that there is no chain from c to d in $N' = N - (a_1 b_1)$. Let M_d consist of d and any node i such that there is a chain from i to d in N', and let $M_c = M - M_d$.

Since $m \geqq 3$ one of these sets has more than one member and neither is empty since $c \varepsilon M_c$ and $d \varepsilon M_d$. If either set, say M_c, has but one member, then the other contains either a_1 or b_1, say b_1. But there is no chain from c to $M_d - b_1$. If both sets have two or more members, remove either a_1 or b_1 and there is no chain between the resulting sets.

For $m \geqq 3$ and $k > 1$ we use an inductive argument. Remove the link $(a_k b_k)$ to obtain N' having degree $k - 1$. By the induction hypothesis there exists a set of no more than $k - 1$ nodes c_j, with $c_j = a_j$ or b_j, $j = 1, 2, \cdots, k - 1$, whose removal from N' results in a disconnected network N''. If either a_k or $b_k \not\varepsilon N''$ we are done; otherwise, call $N'' + (a_k b_k) = N^*$. Since $k < m - 1$, $m^* \geqq m - (k - 1) \geqq 3$, so if N^* is connected we may apply the $k = 1$ case to show that either the removal of a_k or b_k disconnects N^*. Thus, for $k < m - 1$, $\kappa \leqq k$; however, $\kappa \leqq k$ trivially when $k = m - 1$, so we are done.

We now show the right half of the inequality. If $\kappa = m - 1$, it is trivially true, so we suppose $\kappa < m - 1$. Let S be a cut-set of N and call the resulting disconnected network N'. Let $M' = M - S$. M' consists of more than one node since $m - \kappa > 1$, hence there exist $c, d \varepsilon M'$ such that there is no chain from c to d in N'. Let M_c and M_d be defined as in the first half of the proof. One or the other contains no more than half of the nodes of M', i.e., no more than $(m - \kappa + 1)/2$ nodes. Without loss of generality, we may suppose M_c is the smaller set. In N consider the removal of the links (ci) where $i \varepsilon M_c + S$, which are no more than $(m - \kappa + 1)/2 + \kappa - 1 = (m - 1 + \kappa)/2$ in number. Clearly the resulting network is not connected because there is no link for which c is the initial node, which concludes the proof.

The right inequality is weak and may be improved by relating the degree to the diameter of a network. Let δ_{ab} be the shortest chain from a to b in a connected network, then $\delta = \max_{a,b} \delta_{ab}$ is called the *diameter* of the network.

THEOREM 3. *For a connected network of diameter $\delta > 2$ and degree k, $k \leqq (m - \delta)/2 + 1$.*

Proof. If $\delta = m$, then there is a circuit on the nodes of N such that at least one of the nodes is the initial node of only one link, thus $k = 1 = (m - m)/2 + 1$.

Consider $2 < \delta < m$. Let a and b be two nodes having no chain with fewer than δ links from a to b. If $a \neq b$, there are $\delta + 1$ nodes S in the shortest chain from a to b and $m - \delta - 1$ nodes in $M - S$. If $i \varepsilon M - S$

then not both (ai) and $(ib) \varepsilon N$ since $\delta > 2$. Thus, either a is the initial node of no more than $(m - \delta - 1)/2$ links to $M - S$ or b is the end node of no more than $(m - \delta - 1)/2$ links from $M - S$. Furthermore, a is the initial node of only one link to the nodes of S and b is the end node of only one from S, else there is a chain with fewer than δ links from a to b. Consequently, the removal of at most $(m - \delta - 1)/2 + 1 < (m - \delta)/2 + 1$ links disconnects N.

If $a = b$, S has δ nodes and $M - S$ has $m - \delta$, and by a similar argument $k \leqq (m - \delta)/2 + 1$.

Observe that for $\delta > 2$, Theorem 3 implies the right side of Theorem 2, for $k \leqq (m - \delta + 2)/2 \leqq (m - 1)/2 < (m - 1 + \kappa)/2$.

We turn now to Menger's theorem [3]. It is proved for graphs; however, substantially the same proof holds for networks and so we state it in that form: If a network is connected and has no cut node, i. e., index $\kappa \geqq 2$, then from any node a to any node b there are at least two chains which have no intermediate nodes in common. Because of the parallel definitions of degree and index, one is led to inquire if the following analogue to Menger's theorem is true: If a network has degree $k \geqq 2$, then from any node a to any node b there are at least two chains which have no links in common. It is indeed true; one proof parallels very closely the demonstration given by Dirac for a strengthened form of Menger's theorem; cf. [2], p. 72. We shall not include this proof, for the result is included in the following considerably stronger result.

THEOREM 4. *If a network has degree k, then from any node a to any other node b there is a set of at least k chains such that no two have a common link.*

Proof. We proceed by induction on k; for $k = 1$ the theorem is trivial.

If N has degree $k > 1$, select a k-descendant N' of N (i. e., one of the smallest complete k-minimal subnetworks of N, see p. 705 of (A)). It suffices to show the theorem for N'. Let n be the length of the shortest chain from a to b. If $n = 1$, remove the link (ab) yielding a network of degree $k - 1$, which, by the induction hypothesis, has $k - 1$ chains from a to b with no link in more than one of them. But (ab) is not common to any of them, so there are k chains from a to b in N such that no pair has a common link.

The remainder of the argument is an induction on n with k fixed. Let λ be a chain from a to b of length n and let c be the node of λ immediately preceding b. The shortest chain from a to c has $n - 1$ links, so by the

induction hypothesis there exists a set A_1 of k chains from a to c having no link common to any pair. Similarly, there is a set B of k chains from c to b having no link common to any pair. We may suppose that at least one chain of B has a link in common with a chain of A_1, else we are done.

Notation. If g and h are two nodes of a chain λ, let $\lambda(g, h)$ denote the part of λ from g to h.

Suppose $\beta \varepsilon B$ has a link in common with a chain of A_1. Proceed along β opposite to its orientation, i. e., from b toward c, until the first link which is common to a chain, say α, of A_1. Continue further along β until either there is a link common to some $\alpha' \varepsilon A_1$, $\alpha' \neq \alpha$, or until c is reached. Let g be the end node of the common link or c, whichever is appropriate. Observe that α and $\beta(g, b)$ may have several common links. Let h be the first node of α, measured along α from a, such that the links of α and $\beta(g, b)$ for which h is the initial node are different. We call $\beta(h, b)$ the *tail* of β.

The remainder of the proof is concerned with the construction of k chains from a to b which satisfy the conditions of the theorem. Parts of chains in A_1 and B will be used. The construction is expedited by dividing A_1 into a number of classes.

A_1 is given. Suppose A_{j-1}, C_{j-1}, D_{j-1}, E_{j-1}, F_{j-1}, and G_{j-1} to be defined. Then define $A_j = D_{j-1} + E_{j-1}$.

Now, for any $\alpha \varepsilon A_j$, let $\beta^j{}_\alpha$ be the j-th distinct chain of B as measured along α from a, which has a link in common with α. Let $g^j{}_\alpha$ be the first node in α which is initial to a link of $\beta^j{}_\alpha$ which is not also a link of α. Then we define

$$C_j = [\alpha \varepsilon A_j \mid \beta^j{}_\alpha(g^j{}_\alpha, b) \text{ is the tail of } \beta^j{}_\alpha].$$

$$D_j = [\alpha \varepsilon A_j \mid \alpha \varepsilon A_j - C_j, \beta^j{}_\alpha(g^j{}_\alpha, b) \text{ has a link in common with some } \alpha' \varepsilon A_j - C_j, \alpha' \neq \alpha].$$

$$E_j = [\alpha \varepsilon A_j \mid \alpha \varepsilon A_j - C_j, \beta^j{}_\alpha(g^j{}_\alpha, b) \text{ has links in common only with members of } \sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma \text{ and } \beta^j{}_\alpha \text{ is associated with some } \alpha' \varepsilon \sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma, \text{ by its defining property}].$$

$$F_j = [\alpha \varepsilon A_j \mid \alpha \varepsilon A_j - C_j, \beta^j{}_\alpha(g^j{}_\alpha, b) \text{ has links in common only with members of } \sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma, \text{ and } \beta^j{}_\alpha \text{ is not associated with any } \alpha' \varepsilon \sum_{\sigma=1}^{j} C_\sigma + \sum_{\sigma=1}^{j-1} F_\sigma \text{ by its defining property}].$$

$$G_j = [\alpha \varepsilon A_j \mid \text{no } g^j{}_\alpha \text{ exists}].$$

Continue this inductive subdivision of A_1 until $A_\eta \neq 0$ and $A_{\eta+1} = 0$.

Let $W_j = [\alpha(a, g^j{}_\alpha)\beta^j{}_\alpha(g^j{}_\alpha, b) \mid \alpha \varepsilon C_j + F_j]$. As above, we shall speak of the β's as being associated with the corresponding α's according to the definition of W_j. Now suppose $\omega \varepsilon W_i$ and $\omega' \varepsilon W_j$ have a link in common. For simplicity we write $\omega = \alpha\beta$, $\omega' = \alpha'\beta'$, and suppose $i \leqq j$. Either β' has a link in common with α or β with α'. Consider the former case. Certainly $\alpha' \not\varepsilon C_j$ since β' is not a tail, so $\alpha' \varepsilon F_j$. But since β' has a link in common with α, then for some $\rho < i$, $\alpha \varepsilon D_\rho + E_\rho$, which implies that β' has a link in common with some $\alpha^* \varepsilon A_\rho - C_\rho$ or that β' has already been associated with some $\alpha'' \varepsilon \sum_{\sigma=1}^{\rho} C_\sigma + \sum_{\sigma=1}^{\rho-1} F_\sigma$. The latter is impossible since $\alpha' \varepsilon F_j$. In the former, continue along β' toward b; there is a last chain $\lambda \varepsilon A_\rho - C_\rho$ which has a link in common with β'. Either $\lambda \varepsilon C_\rho + F_\rho$ or β' has already been associated with a member of A_1, both of which are impossible. So β' and α do not have a common link. For the second case, in which β has a link in common with α', we may suppose $i < j$ since the case $i = j$ has already been covered. Thus, $\alpha \varepsilon F_i$, but since $\alpha' \varepsilon A_j \subset A_i - C_i$ and β has a link in common with α', then $\alpha \varepsilon D_i$ by definition. This is a contradiction.

Let $W' = \sum_{j=1}^{\eta} W_j$ have r members; we have shown there are r chains from a to b such that no two have a link in common.

Let $G = \sum_{j=1}^{\eta} G_j$ have s members, then we show $s + r = k$. If S is a finite set we denote by $N(S)$ the number of elements in S. By definition

$$A_j = C_j + F_j + G_j + A_{j+1},$$

and since the sets on the right are mutually exclusive,

$$\sum_{j=1}^{\eta} N(A_j) = \sum_{j=1}^{\eta} [N(C_j) + N(F_j)] + \sum_{j=1}^{\eta} N(G_j) + \sum_{j=2}^{\eta+1} N(A_j).$$

By choice, $N(A_{\eta+1}) = 0$ and $N(W_j) = N(C_j) + N(F_j)$, so

$$N(A_1) = k = \sum_{j=1}^{\eta} N(W_j) + \sum_{j=1}^{\eta} N(G_j).$$

But, $W_i \cap W_j = 0$, $G_i \cap G_j = 0$, so $N(W') = \sum_{j=1}^{\eta} N(W_j)$ and $N(G) = \sum_{j=1}^{\eta} N(G_j)$, so $k = r + s$.

Let B' be the set of $\beta \varepsilon B$ which have no link in common with any $\omega \varepsilon W'$. Clearly, $N(B') = N(A_1) - N(W') = k - r = s$. Thus, we may set up an arbitrary $1:1$ relation between the s elements of G and the s elements of B'.

Denote it by a subscript q and call the set of chains $\alpha_q\beta_q$, $q = 1, 2, \cdots, s$, $\alpha_q \varepsilon G$, $\beta_q \varepsilon B'$, W''. $W = W' + W''$ is a set of k chains from a to b, which we now show concludes the proof.

First, let $\omega' = \alpha'\beta' \varepsilon W'$ and $\omega'' = \alpha''\beta'' \varepsilon W''$. By definition of W'', β'' has no link in common with α'. If β' has a link in common with α'', then since $\alpha'' \varepsilon G$, β' must have been associated with some $\alpha^* \neq \alpha'$, whence $\alpha'\beta' \not\varepsilon W'$. Finally, if $\omega = \alpha\beta$ and $\omega' = \alpha'\beta' \varepsilon W''$ and, say, β has a link in comon with α', then since $\alpha' \not\varepsilon \sum_{\sigma=1}^{\eta} C_\sigma + \sum_{\sigma=1}^{\eta} F_\sigma$, β must have been associated with some α^*, whence $\beta \not\varepsilon B'$. The proof is concluded.

From the analogue to Menger's theorem one may deduce the structure of 2-minimal graphs.

THEOREM 5. *If G is a 2-minimal graph, there exist minimal subnetworks N_1 and N_2 such that $G = N_1 + N_2$ and N_1 and N_2 are (opposite) orientations of G.*

Proof. Let N be a descendant of G and suppose there is an arc $ab \varepsilon G - N$. Let α be a chain of N from a to b and β from b to a. There is a link (cd) of α such that (dc) is in β, for otherwise there are two chains of arcs between a and b in $G - ab$ having only nodes in common. Thus $G - (ab)$ has degree $\geqq 2$, which is contrary to assumption. Let $N' = N + (ab) - (cd)$. If N' is connected it is also minimal since it has the same number of links as N. To show it connected it is sufficient to show a chain from b to a and one from c to d. The chain β from b to a remains and $\beta(c, a)(ab)\beta(b, d)$ exists.

If $G - N'$ has an arc, continue the process until N_1 is obtained such that $G - N_1 = N_2$ is arc-free. N_1 is also arc-free, for if $ab \varepsilon N_1$ then by Theorem 3. 4 of (A) N_1 consists of two disjoint connected subnetworks joined only by ab. But since G is 2-minimal there is another chain of arcs from a to b not including ab, so N_2 has an arc, a contradiction. Since N_1 is arc-free it is an orientation of G, hence N_2 is a connected orientation of G, and so is minimal.

Finally, it should be observed that Theorem 4, a generalization of a result suggested by Menger's theorem, in turn suggests a generalization to his theorem, to wit: *If a network has index κ, then from any node a to any other node b there exists a set of at least κ chains such that no two have a common intermediate node.* Since the proof of Theorem 4 is based on two sets of chains with a common node c, it is evident that no minor modification of that proof will suffice to demonstrate the above statement, and I have been unable to develop a proof of it.

Some interest attaches in either proving it or giving a counter example, for if it is true there are theorems in graph theory (cf. [2], Theorems 1, 4, 5) of the form "If a graph has no cut-node, then . . ." which presumedly can be strengthened to a form "If a graph has index κ, then . . ."

4. *h*-transitive networks. As was pointed out in the introduction, the conditions of transitivity and connectedness result in the single class of networks, the complete graphs, so it is desirable to weaken the transitivity condition. We shall call a network N *h-transitive* if there is at least one chain $(ab, h) \varepsilon N$ such that $a \neq b$, and if for every chain (cd, q) such that $c \neq d$, then $(cd) \varepsilon N$ if $q \geqq h + 1$ and $(cd) \not\varepsilon N$ if $1 < q \leqq h$. Clearly, $1 \leqq h \leqq m - 2$, and for connected networks, 1-transitivity implies transitivity.

For connected networks, two cases can be distinguished: either there exists a chain of length $\geqq h + 1$, or there does not. In the latter case, it is easy to see that the network is minimal. This case has been discussed in (A), so we shall be interested only in the former case.

The following are a set of examples of non-minimal, h-transitive networks with $m \geqq h + 2 \geqq 5$. Let Q be a set of four nodes 1, 2, 3, and 4, R a set of $h - 2$ nodes distinct from Q labeled $5, 6, \cdots, h + 2$, and S a set of $m - h - 2$ nodes disjoint from $Q + R$. Let the following links be present on $Q + R + S$: (13), (14), (23), (24), (35), (45), (56), $\cdots$, $(h + 1, h + 2)$, $(h + 2, i)$, $(i3)$, $(i4)$, where $i \varepsilon S$. It is not difficult to show these networks satisfy the above requirements.

LEMMA 2. *If N is h-transitive, $h > 1$, and there exists $(ab, q) \varepsilon N$ with $q > h$, then $q = h + 1$.*

Proof. $(ab, q) = (ac)(cb, q - 1)$, and if $q > h + 1$, $q - 1 > h$, so (cb), $(ab) \varepsilon N$. But for $h \geqq 2$, $(ac)(cb) \varepsilon N$ implies $(ab) \not\varepsilon N$, a contradiction.

In (A) a network was called *uniform* if every connected subnetwork has degree 1. A graph which consists of only a circuit of arcs encompassing all the nodes is called a *circle*.

THEOREM 6. *If a network is connected and h-transitive, $h > 1$, then it is uniform or a circle (which is 2-minimal and for which $h = m - 2$).*

Proof. If N is minimal, it is uniform (p. 704 of (A)).

If N is non-minimal, there is an $h + 1$ chain and, by Lemma 2, it is the longest chain in N. Let its nodes be ordered by the orientation of the chain and $M_1 = \{a, a + 1, \cdots a + h, a + h + 1 = b\}$ and $M_2 = M - M_1$.

If $h < m - 2$, then $M_2 \neq 0$. If $(ba) \varepsilon N$, then a simple induction shows there is a circle on M_1. Then, any link from a node of M_1 to one of M_2 results in an $h + 2$-chain, and at least one such link exists since N is connected. By Lemma 2, this is impossible, so $(ba) \not\varepsilon N$. Let $c \varepsilon M_2$, then $(bc) \not\varepsilon N$ or $(ab, h + 1)(bc)$ would be an $h + 2$ chain. But since N is connected, there exists at least one $a + i \varepsilon M_1$, $1 \leqq i \leqq h$, such that $(b, a + i) \varepsilon N$. However, for $j > i$, $(b, a + j) \not\varepsilon N$ since $(b, a + i)(a + i, a + i + 1) \cdots (a + j - 1, a + j)$ is a chain of length no greater than $1 + (h - 1) = h$. Therefore, b is the initial node of exactly one link, so N has degree 1.

If $h = m - 2$, then $M_2 = 0$. The only possible links to the node $a + h$ are $(b, a + h)$ and $(a + h - 1, a + h)$, since any others produce a chain (ab, q) with $q \leqq h$. Thus, the degree of N is, in this case, no greater than 2.

Suppose N is $(m - 2)$-transitive and of degree 2, then we show N is a circle (the converse is trivial). Since $k = 2$, $(b, a + h) \varepsilon N$. Now node $a + h - 1$ must be the end node of at least two links, one being $(a + h - 2, a + h - 1)$. Of the other two possibilities, $(b, a + h - 1)$ and $(a + h, a + h - 1)$, the former is excluded because $(b, a + h - 1)(a + h - 1, a + h)$ implies $(b, a + h) \not\varepsilon N$, contrary to what we have just shown. Proceed inductively and a circle results.

Now consider the non-minimal h-transitive networks of degree 1. Let S be a connected subnetwork of N, and let h' be the length of the longest chain in S. Either $h' = h + 1$ or $h' \leqq h$. In either case, S is h'-transitive, and so the degree of S is 1 except, possibly, if $h' = 1$ or $h' = m' - 2$. If $h' = 1$, then since $h > 1$, $m' = 2$, and so the degree is 1. If $h' = m' - 2$, the only interesting case is degree 2, which, by what we have just seen, implies S is a circle. But, then, $h' = h$, and N is a circle, for $h = m - 2$, else there is an $h + 2$ chain. This is contrary to assumption, so S has degree 1, and N is uniform.

The second example on p. 719 of (A) shows there are uniform networks which are not h-transitive.

COROLLARY. *For $m \geqq 5$, there are no 2-transitive, connected, non-minimal networks.*

Proof. Suppose N is 2-transitive, connected, and non-minimal. Let the nodes of one of the 3-chains be a, $a + 1$, $a + 2$, b. As in the first part of the above proof, if $m \geqq 5$, there is a link from b to $a + i$, $1 \leqq i \leqq 2$. If $(b, a + 1) \varepsilon N$, then $(ab)(b, a + 1)(a + 1, a + 2) \varepsilon N$ implies $(a, a + 2) \varepsilon N$, which is impossible. Thus, for N to be connected, $(b, a + 2) \varepsilon N$. If $(a + 2, a) \not\varepsilon N$, then there is a 3-chain from b to a, which is impossible.

But, $(b, a+2)(a+2, a)(a, a+1) \varepsilon N$ implies $(b, a+1) \varepsilon N$, which we have just shown is impossible. Thus, N does not exist.

THEOREM 7. *Let N be connected, non-minimal, h-transitive, $h > 1$, and not a circle. If N contains an arc ab, then N consists of two connected subnetworks N_a and N_b joined only by the arc ab. Either N_a or N_b is h-transitive and non-minimal, and the other is either minimal or h-transitive and non-minimal.*

Proof. Let
$$M_a = [i \varepsilon M \mid \text{there exists } (ai, q) \text{ not including } b] + a,$$
$$M_b = [i \varepsilon M \mid \text{there exists } (bi, q) \text{ not including } a] + b,$$
$$M_{ab} = M_a \cap M_b,$$
and
$$M'_a = [i \varepsilon M \mid \text{there exists } (ia, q) \text{ not including } b] + a,$$
$$M'_b = [i \varepsilon M \mid \text{there exists } (ib, q) \text{ not including } a] + b,$$
$$M'_{ab} = M'_a \cap M'_b.$$

Observe that $M = M_a + M_b = M'_a + M'_b$. If $i \varepsilon M_a \cap (M - M'_a)$, then there exist $(ai, r) \varepsilon N$ not containing b and all $(ia, s) \varepsilon N$ do contain b, so there exists a chain α from a to b not including (ab). Since $(ab) \varepsilon N$, which is h-transitive, α must be of length $h+1$. Since $(ba) \varepsilon N$, we may use the same induction as in the proof of Theorem 6 to show there is a circle of arcs on the nodes of α. This contradicts the fact that N is uniform. Thus, $M_a \subset M'_a$. Similarly, $M'_a \subset M_a$, so $M_a = M'_a$. In like manner, $M_b = M'_b$. Thus, $M_{ab} = M_a \cap M_b = M_a \cap M'_b$, so by the same argument $M_{ab} = 0$. Similarly, $M'_{ab} = 0$. Let N_a and N_b be the maximal subnetworks of N on M_a and M_b. By what we have just shown they are connected and they have no node in common. They are joined by ab, and no other link exists between them since $M_{ab} = 0$.

Not both N_a and N_b are minimal, for if they were then N would be minimal. Indeed, no $h+1$ chain traverses the arc ab, for if it did, there would exist another link between N_a and N_b. Thus, one of them is h-transitive and non-minimal, and the other is minimal or h-transitive, non-minimal.

The class of non-minimal, h-transitive, uniform networks on m nodes is smaller than the class of minimal networks on m nodes, and the former can be readily obtained from the latter. Observe, if N is h-transitive and non-minimal, it contains a minimal N' as a descendant. N' is $h+1$-transitive, and N is obtained inductively from N' by introducing a link (ab) every time a chain $(ab, h+1)$ appears. It is easy to find examples of

minimal networks for which this operation does not result in an h-transitive network, so the class of minimal networks is the larger.

For example, if $m = 5$, it is easy to construct the 15 possible minimal networks using Theorem 3. 4 of (A). Of these, 10 have arcs and in each case the longest chain in the network passes through the arc, so by Theorem 7 they cannot be descendants of an h-transitive non-minimal network. Of the remaining five, one is the circuit which obviously becomes the circle, and one has $h = 3$ which by the corollary to Theorem 3 cannot yield a 2-transitive case. Performing the inductive operation described above on the other three gives the complete graph in two cases and a 3-transitive network in the third case (which is included in the example at the beginning of this section).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] G. A. Dirac, "Note on the colouring of graphs," *Mathematische Zeitschrift*, vol. 54 (1951), pp. 347-353.

[2] ———, "Some theorems on abstract graphs," *Proceedings of the London Mathematical Society*, ser. 3, vol. 2 (1952), pp. 69-81.

[3] D. Konig, *Theorie der endlichen und unendlichen Graphen*, New York, Chelsea, 1950.

[4] R. D. Luce, "Two decomposition theorems for a class of finite oriented graphs," *American Journal of Mathematics*, vol. 74 (1952), pp. 701-722.

MODULES OVER OPERATOR ALGEBRAS.*[1]

By Irving Kaplansky.

1. Introduction. In [3], Th. 3 the author proved that any *-automorphism of an AW^*-algebra[2] of type I, leaving the center elementwise fixed, is inner. Now in the case of a factor (that is, the algebra of all bounded operators on a Hilbert space) a better result is known, for then *every* automorphism is inner. This leaves a gap that deserves to be filled. A companion problem is the following: is every derivation of an AW^*-algebra of type I inner? It appears that existing AW^* techniques are inadequate to solve these problems, and this paper is devoted to introducing a new technique that does the trick. In brief: the new idea is to generalize Hilbert space by allowing the inner product to take values in a more general ring than the complex numbers. After the appropriate preliminary theory of these AW^*-modules has been developed, one can operate with a general AW^*-algebra of type I in almost the same way as with a factor.

Besides solving the two problems mentioned above, the introduction of AW^*-modules simplifies portions of [3], and also enables us to settle the existence question left open there: we are now able to construct an $\aleph$-homogeneous AW^*-algebra of type I for any prescribed $\aleph$ and center.

In the more special case of W^*-algebras, the two problems could be handled by available tools. But even here AW^*-modules seem to provide the natural method. One may expect that the theory of AW^*-modules will have further applications, both to W^* and to AW^*-algebras.

2. C^*-modules. Let A be a commutative C^*-algebra with unit,[3] and let H be an A-module in the ordinary algebraic sense (including the assumption that the unit element of A acts as unit operator). We shall put the elements of A (typically $a, b, \cdots$) on the left of the elements of H (typically

* Received May 12, 1953.

[1] This paper was prepared with the partial support of the Office of Naval Research.

[2] Definitions will not be repeated from [2] and [3]; but few actual results will be used from the two papers.

[3] The assumption of a unit element is not vital, but it seems pointless to omit it, since A will shortly be an AW^*-algebra. On the other hand, extension of the theory to modules over non-commutative C^*-algebras presents many difficulties.

$x, y, \cdots$). Suppose there is defined on H an inner product taking values in A and satisfying:

(1) $(x, y) = (y, x)^*$,

(2) $(x, x) \geqq 0$ and is 0 only for $x = 0$,

(3) $(ax + x_1, y) = a(x, y) + (x_1, y)$,

for all x, x_1, y in H and a in A. Of the immediate consequences of the axioms we call explicit attention only to the following: $(x, ay) = a^*(x, y)$, and we pass on at once to the introduction of a norm. As compared with Hilbert space, there are in fact two "norms" available, one A-valued and the other numerical; they both have a role to play in our work. We use the notation

$$| x | = (x, x)^{\frac{1}{2}}, \qquad \| x \| = \| (x, x) \|^{\frac{1}{2}},$$

where on the right we mean the usual positive square root and norm in A. We have that $\| x \|$ is the norm of $| x |$ in the algebra A; alternatively, $\| x \|$ is the sup of $| x |$ when the latter is regarded as a function on the space of maximal ideals in A.

The Schwarz inequality

$$(1) \qquad | (x, y) | \leqq | x | \, | y |$$

can be verified by adapting a standard proof. It is also possible to reduce to the numerical case by the following device. For any maximal ideal M in A we may define a numerical inner product $(x, y)(M)$ on H, by mapping modulo M. In this inner product there may exist non-zero elements x with $(x, x)(M) = 0$; nevertheless the Schwarz inequality

$$(2) \qquad | (x, y))(M) | \leqq | x | (M) | y | (M)$$

is known to hold. The validity of (2) for every M is precisely equivalent to (1). On taking norms in (1) we further get the numerical version of the Schwarz inequality:

$$(3) \qquad \| (x, y) \| \leqq \| x \| \, \| y \|.$$

From (1) or (3) we deduce in the usual way the triangle inequality $\| x + y \| \leqq \| x \| + \| y \|$, and we have that H is a normed linear space. By (3) the inner product is jointly continuous in its arguments; likewise the equality $| ax | = | a | \, | x |$ shows that ax is jointly continuous in its arguments. So H may be completed with preservation of all the postulates. If H is already complete, we shall call it a *C^*-module over A*. Thus a C^*-

module is a kind of blend of a Hilbert space and a commutative C^*-algebra.

By a *bounded operator* T from a C^*-module H into a second one K we mean a mapping of H into K which is not only linear and continuous as usual, but also a *module homomorphism.* Thus (if we place T on the right of the elements of H) we have $(ax)T = a(xT)$ for all a and x. The set B of all bounded operators on H forms a Banach algebra in the usual operator norm.

We call T^* the *adjoint* of T if $(xT, y) = (x, yT^*)$ for all x and y. The question of the existence of T^* will be discussed in § 8 only with the aid of more axioms. But whenever T^* exists, we can verify the equation $\| TT^* \| = \| T \|^2$.

LEMMA 1. *Let H be a C^*-module, and T a bounded operator on H with an adjoint T^* which is also a bounded operator. Then $\| T \| = \| T^* \|$ and $\| TT^* \| = \| T \|^2$.*

Proof. We have

$$| xT |^2 = (xT, xT) = (xTT^*, x) \leqq | x | \, | xTT^* |, \tag{4}$$

the last step by the Schwarz inequality (1). We take norms in (4), recalling that $\| T \|$ is the sup of $\| xT \|$ for $\| x \| \leqq 1$. The result is the first inequality in

$$\| T \|^2 \leqq \| TT^* \| \leqq \| T \| \, \| T^* \|, \tag{5}$$

while the second inequality holds in any Banach algebra. On cancelling $\| T \|$ we get $\| T \| \leqq \| T^* \|$. Since T is also the adjoint of T^*, the reverse inequality also holds. Hence $\| T \| = \| T^* \|$, and insertion of this into (5) completes the proof.

3. AW^*-modules. It does not appear to be possible to get much deeper into the subject without imposing further postulates, acting as a sort of algebraic substitute for a weak topology. These postulates are motivated by two properties of an AW^*-algebra A ([2], Lemmas 2. 2 and 2. 5).

(a) Let $\{e_i\}$ be orthogonal projections in A with l. u. b. e, and suppose a is an element of A with $e_i a = 0$ for all i; then $ea = 0$.

(b) Let $\{e_i\}$ be central orthogonal projections in A with l. u. b. 1, and let $\{a_i\}$ be a bounded subset of A; then there exists in A an element a with $e_i a = e_i a_i$ for all i.

We propose to assume outright the analogues of these two properties.

12

Definition. Let A be a commutative AW^*-algebra. We say that H is an *AW^*-module over A* if it is a C^*-module over A and further enjoys the following two properties:

(a) Let $\{e_i\}$ be orthogonal projections in A with l. u. b. e, and suppose x is an element of H with $e_i x = 0$ for all i; then $ex = 0$.

(b) Let $\{e_i\}$ be orthogonal projections in A with l. u. b. 1, and let $\{x_i\}$ be a bounded subset of H; then there exists in H an element x with $e_i x = e_i x_i$ for all i.

It follows from postulate (a) that the element x of (b) is unique, and we shall write $x = \sum e_i x_i$. This is in accordance with the notation $\sum e_i a_i$ used in [3] for the analogous element constructed in the algebra A. These two infinite "sums" are related in the desired way.

LEMMA 2. *Let A be a commutative AW^*-algebra, and $\{e_i\}$ a set of orthogonal projections in A with* l. u. b. 1. *Let H be an AW^*-module over A, y an element of H, and $\{x_i\}$ a bounded subset of H. Then*

$$(\sum e_i x_i, y) = \sum e_i (x_i, y). \tag{6}$$

Proof. It is to be observed that the elements (x_i, y) are bounded in A, and so the right side of (6) is well defined. To prove (6) it is enough to verify that it holds after multiplication by a fixed e_j (we shall be making repeated use of this observation throughout the paper). Then the right side becomes $e_j(x_j, y)$, while the left side is

$$e_j(\sum e_i x_i, y) = (e_j \sum e_i x_i, y) = (e_j x_j, y) = e_j(x_j, y).$$

LEMMA 3. *Let H be an AW^*-module over A. Then the annihilator in A of any subset of H is a direct summand of A.*

Proof. Let I be the annihilator in question; I is an ideal by algebra, and is closed in the norm of A by the continuity of the module operations. Let $\{e_i\}$ be a maximal set of orthogonal projections in I. By postulate (a), their l. u. b. e is again in I. We claim that $I = eA$, and it suffices to prove $I = eI$, or equivalently $(1 - e)I = 0$. If on the contrary $(1 - e)I$ is non-zero, it contains a non-zero projection, which could be used to enlarge the set $\{e_i\}$.

In particular, the annihilator of all of H is a direct summand of A, and we say that H is *faithful* if its annihilator is 0. As a rule, there is no loss of generality in restricting our attention to faithful AW^*-modules.

LEMMA 4. *Let x be a non-zero element of an AW^*-module H over A. Then there exists in A a non-zero projection e and an element a with $a \mid x \mid = e$.*

Proof. We have that $\mid x \mid$ is a non-zero element of the commutative AW^*-algebra A. Consequently there exists a non-zero projection e such that the function representing $\mid x \mid$ is bounded away from zero in the direct summand eA. We take a to be the inverse of $\mid x \mid$ in eA.

LEMMA 5. *A faithful AW^*-module H contains an element x with $\mid x \mid = 1$.*

Proof. Lemma 4 provides us with an element x_1 with $\mid x_1 \mid$ equal to a non-zero projection e_1. We apply Zorn to get a maximal collection $\{x_i\}$ with $\mid x_i \mid = e_i$, and $\{e_i\}$ orthogonal projections, say with l. u. b. e. If $e \neq 1$, we have $(1 - e)H \neq 0$ since H is faithful. Applying Lemma 4 to a non-zero element of $(1 - e)H$, we enlarge the collection $\{x_i\}$. Hence e must be 1. Since each $\mid x_i \mid = 1$, we may form the element $x = \sum e_i x_i$. For any e_j we have

$$e_j(x, x) = (e_j x, e_j x) = (e_j x_j, e_j x_j) = e_j.$$

Thus $(x, x) - 1$ annihilates every e_j and is 0. Hence x is the desired element.

It is appropriate to observe that Lemma 5 may fail if H is merely a C^*-module (even over an AW^*-algebra). For instance take A to be the algebra of all bounded sequences of complex numbers, and H the C^*-module of all sequences aproaching 0.

We now introduce some terminology, imitating the example of Hilbert space. We say that x and y are *orthogonal if* $(x, y) = 0$. The *orthogonal complement* R' of a subset R of H is the set of all x with $(R, x) = 0$. A set $\{x_\lambda\}$ is *orthonormal* if each $\mid x_\lambda \mid = 1$ and $(x_\lambda, x_\mu) = 0$ for $\lambda \neq \mu$. The existence of a maximal orthonormal set is assured by Zorn's lemma. But this is not what we generally need. Rather we want what we shall call an orthonormal basis.

Definition. An *orthonormal basis* in an AW^*-module is an orthonormal set whose orthogonal complement is 0. An AW^*-module is said to be *homogeneous* if it possesses an orthonormal basis, or more precisely it is $\aleph$-homogeneous if it possesses an orthonormal basis of $\aleph$ elements.[4]

[4] This terminology is not meant to suggest that the $\aleph$ in question is unique. As regards this problem, we have not advanced beyond what was shown in [3]: $\aleph$ is unique

Not every AW^*-module has an orthonormal basis, but we can split it into homogeneous parts which do. To accomplish this we next launch the theory of submodules.

4. Submodules. The appropriate concept for our purposes is embodied in the following definition.

Definition. Let H be an AW^*-module over A. By an *AW^*-submodule* S we mean a subset satisfying: (1) S is a submodule of H in the ordinary algebraic sense, (2) S is closed in the norm topology of H, (3) if $\{x_i\}$ is a bounded subset of S, and $\{e_i\}$ are orthogonal projections in A with l. u. b. 1, then the element $\sum e_i x_i$ is again in S.

We remark: (a) an AW^*-submodule is itself an AW^*-module over A, (b) the intersection of any number of AW^*-submodules is again an AW^*-submodule, (c) consequently for any subset R there exists a smallest AW^*-submodule containing R; we call it the AW^*-submodule generated by R.

The next two lemmas provide us with two natural sources of AW^*-submodules.

LEMMA 6. *Let H be an AW^*-module. Then the orthogonal complement of any subset of H is an AW^*-submodule of H.*

Proof. The only point worthy of note is this: given that $(x_i, y) = 0$ for every i, prove that $(\sum e_i x_i, y) = 0$. This follows at once from Lemma 2.

LEMMA 7. *Let T be a bounded operator from an AW^*-module H into a second one. Then the kernel N of T is an AW^*-submodule of H.*

Proof. We content ourselves with proving that $x_i \in N$ implies $\sum e_i x_i \in N$. It is enough to prove $(\sum e_i x_i)T = 0$ after multiplication by a fixed e_j, after which it becomes $(e_j x_j)T$, which does vanish.

Now consider a faithful AW^*-module H, a maximal orthonormal set $\{x_\lambda\}$ in H, and its orthogonal complement S. By Lemma 6, S is an AW^*-submodule of H. If S is faithful we can use Lemma 5 to enlarge the set $\{x_\lambda\}$. Hence there must exist a non-zero projection e in A with $eS = 0$. We now claim that the elements $\{ex_\lambda\}$ constitute an orthonormal basis for eH, regarded

---

if A satisfies the countable chain condition locally. The author conjectures that the uniqueness may fail otherwise.

as an AW^*-module over eA. For suppose the element y in eH is orthogonal to all ex_λ. Then

$$0 = (ex_\lambda, y) = e(x_\lambda, y) = (x_\lambda, ey).$$

Hence $ey = y$ is in S, and this implies $y = 0$, since $eS = 0$. When pursued by transfinite induction, the process yields the following theorem.

THEOREM 1. *Let H be a faithful AW^*-module over A. Then there exist in A orthogonal projections $\{e_i\}$ with* l. u. b. 1 *such that each e_iH is a homogeneous AW^*-module over e_iA.*

5. Construction of AW^*-modules. It is time for us to consider the question of the existence of AW^*-modules. Now our fundamental example is A itself, regarded as a module over A. There is no difficulty in discussing the direct sum of a finite number of copies of A. But the construction of an "infinite direct sum" requires more elaborate discussion. While we are at it, we might as well construct the direct sum of arbitrary AW^*-modules over A.

Our fundamental tool for this purpose is the use of the infinite sums in A which were briefly discussed in § 4 of [3]. We recall that the self-adjoint elements of a commutative AW^*-algebra A form a conditionally complete lattice, that is, every bounded set has a least upper bound. If elements $a_\lambda \geqq 0$ are given in A such that there is a fixed upper bound to all finite sums, we say that $\sum a_\lambda$ converges and we define the sum to be the least upper bound of these finite sums. There should be no danger of confusion between these sums and the other kind of infinite sum $\sum e_i x_i$ which we are also using; in particular we shall always use Greek subscripts for the former and Latin for the latter.

The reader should be warned of a possible pitfall: if we think of a_λ as a function on the space of maximal ideals of A, then $\sum a_\lambda$ is not the point-wise sum of these functions (although the two sums differ only on a set of the first category).

We shall further need to make use of sums of elements which are not necessarily self-adjoint or positive. Only absolute convergence is relevant, and so we define $\sum a_\lambda$ to be convergent if $\sum | a_\lambda |$ is. The actual value of the sum $\sum a_\lambda$ is assigned by splitting a_λ into four parts (first into real and imaginary parts, then each of these into positive and negative parts). We shall not pause over routine facts needed in manipulating these sums, samples of which are

$$| \sum a_\lambda | \leqq \sum | a_\lambda |, \qquad b \sum a_\lambda = \sum b a_\lambda.$$

However a crucial rôle is played by a suitable Schwarz inequality.

LEMMA 8. *If* $\sum | a_\lambda |^2$ *and* $\sum | b_\lambda |^2$ *converge, so does* $\sum a_\lambda b_\lambda$, *and*

$$| \sum a_\lambda b_\lambda |^2 \leqq \sum | a_\lambda |^2 \sum | b_\lambda |^2. \tag{7}$$

Proof. We first prove (7) for finite sums, by adapting a standard proof, or by arguing modulo maximal ideals of A. Then, holding a finite sum fixed on the left of (7), we may put in the infinite sums on the right. Since the resulting inequality holds for every finite sum on the left of (7), we get the convergence of $\sum a_\lambda b_\lambda$ and the desired inequality.

Now let an index set I be given, and for each $\lambda \in I$ an AW^*-module H_λ over A. Define H to be the set of all arrays $x = \{x_\lambda\}$ with $x_\lambda \in H_\lambda$, and $\sum | x_\lambda |^2$ convergent. For $x = \{x_\lambda\}$ and $y = \{y_\lambda\}$ in H define (x, y) to be $\sum (x_\lambda, y_\lambda)$. Since $|(x_\lambda, y_\lambda)| \leqq | x_\lambda | \, | y_\lambda |$, the convergence of $\sum (x_\lambda, y_\lambda)$ is assured by Lemma 8. We pass rapidly over the fact that H is in a natural way a C^*-module over A, the verification being routine except for the completeness of H in its norm. This we prove in (c) below, after we have checked the two AW^* postulates.

(a) Let there be given $x = \{x_\lambda\}$ in H, and orthogonal projections e_i in A with l. u. b. e. Suppose each $e_i x = 0$. We have to prove $ex = 0$. Now $e_i x_\lambda = 0$, whence $e x_\lambda = 0$, $ex = 0$.

(b) Let there be given elements $x(i) = \{x_\lambda(i)\}$ in H, with $\| x(i) \|^2$ bounded by a constant K. (Here i of course runs over a second independent index set). Let e_i be orthogonal projections in A with l. u. b. 1. We must construct an element y in H with $e_i y = e_i x_i$ for all i. Now for fixed λ we have $| x_\lambda(i) |^2 \leqq K$. Hence in the AW^*-module H_λ we may form the element $y_\lambda = \sum e_i x_\lambda(i)$, satisfying $e_i y_\lambda = e_i x_\lambda(i)$ for all i. When λ is restricted to a finite subset J of the index set I, we have

$$\sum_{\lambda \in J} | y_\lambda |^2 = \sum_{\lambda \in J} \sum_i e_i | x_\lambda(i) |^2. \tag{8}$$

When (8) is multiplied by e_j, the right side becomes

$$\sum_{\lambda \in J} e_j | x_\lambda(j) |^2$$

and is bounded by $\| x(j) \|^2 \leqq K$. Since this is true for every j, the left side of (8) is likewise bounded by K. Hence $y = \{y_\lambda\}$ is in H. Moreover

$$e_i y = \{e_i y_\lambda\} = \{e_i x_\lambda(i)\} = e_i x_i,$$

and y is the desired element.

(c) Let $x(m) = \{x_\lambda(m)\}$, with m running over the positive integers, be elements constituting a Cauchy sequence in H (relative to the norm in H). We shall produce in H a limit y for this sequence. We begin by noting that for each fixed λ, the elements $x_\lambda(m)$ form a Cauchy sequence in H_λ, converging say to y_λ. The numbers $\| x(m) \|^2$ are bounded, say by K. We claim that also $\sum | y_\lambda |^2 \leqq K$. It is enough to verify this for a finite sum, taken over a finite subset J of the index set I. Now

$$\sum_{\lambda \varepsilon J} | y_\lambda |^2 = \lim_{m \to \infty} \sum_{\lambda \varepsilon J} | x_\lambda(m) |^2,$$

while for each m

$$\sum_{\lambda \varepsilon J} | x_\lambda(m) |^2 \leqq \| x(m) \|^2 \leqq K.$$

Hence $\sum \; y_\lambda |^2 \leqq K$, which means that $y = \{y_\lambda\}$ is in H. It remains to prove that $x(m)$ converges to y in the norm of H. Given $\epsilon > 0$, cut in far enough so that all $\| x(m) - x(n) \|^2 \leqq \epsilon$. We claim that

$$\sum_{\lambda} | x_\lambda(m) - y |^2 \leqq \epsilon.$$

For (again) it is enough to verify this with λ restricted to a finite subset J. But

$$\sum_{\lambda \varepsilon J} | x_\lambda(m) - y_\lambda |^2 = \lim_{n \to \infty} \sum_{\lambda \varepsilon J} | x_\lambda(m) - x_\lambda(n) |^2 \leqq \epsilon.$$

Hence $\| x(m) - y \|^2 \leqq \epsilon$, and this proves the convergence of $x(m)$ to y.

We have thus completed the proof that H is an AW^*-module. Let us now specialize to the case where each H_λ is isomorphic to A. Then H has an evident orthonormal basis: the elements which are 1 at a designated coordinate λ and 0 everywhere else. The cardinal number of this orthonormal basis is the same as the cardinal number of the index set I, and this is at our disposal. We have proved:

THEOREM 2. *For any commutative AW^*-algebra A and cardinal number $\aleph$, there exists an $\aleph$-homogeneous AW^*-module over A.*

6. Orthogonal decomposition. We wish now to prove that any AW^*-submodule is a direct summand, and also that a homogeneous AW^*-module is determined in a suitable way by an orthonormal basis. The following lemma provides the basic information needed in proving both of these facts.

LEMMA 9. *Let H be an AW^*-module over A, let $\{x_\lambda\}$ be an orthonormal set in H, let S be the AW^*-submodule of H generated by $\{x_\lambda\}$, and let $\{c_\lambda\}$*

be elements in A such that $\sum | c_\lambda |^2$ is convergent. Then there exists in S an element t satisfying $| t |^2 = \sum | c_\lambda |^2$, $(t, x_\lambda) = c_\lambda$ for all λ.

Proof. The proof is very much the same as that of Lemma 7 in [3], and consequently we shall not give full details.

Write $w_\lambda = | c_\lambda |^2$, $w = \sum w_\lambda$. Let an integer m be given. We apply [3], Lemma 5, to obtain a set $\{e_i\}$ of orthogonal projections in A with l. u. b. 1, and for each i a certain finite sum v_i of w_λ's such that

$$\| e_i(w - v_i) \| < m^{-2}. \tag{9}$$

Write u_i for $\sum c_\lambda x_\lambda$, taken over the same finite set of λ's as were used in forming v_i. We have $| u_i | \leqq w$, and so we may define t_m in H by $t_m = \sum e_i u_i$. The proof that $\{t_m\}$ is a Cauchy sequence (in the norm topology for H) does not differ materially from the corresponding portion of the proof of [3], Lemma 7, and we omit it. Let t denote the limit of t_m as $m \to \infty$. We note that $u_i \in S$, $t_m \in S$, and so $t \in S$.

It remains to prove that t has the two properties claimed for it in the lemma. We consider (u_i, x_λ) and note that it is c_λ, if λ is one of the subscripts occurring in the defining sum $u_i = \sum c_\lambda x_\lambda$, and is 0 otherwise. In the latter case $e_i | c_\lambda |^2$ is bounded by $e_i(w - v_i)$, and so $| e_i c_\lambda | < 1/m$ by (9). Also

$$(t_m, x_\lambda) = (\sum e_i u_i, x_\lambda) = \sum e_i (u_i, x_\lambda)$$

by Lemma 2. From this it follows that

$$| (t_m, x_\lambda) - c_\lambda | < 1/m$$

always holds. Proceeding to the limit as $m \to \infty$, we deduce $(t, x_\lambda) = c_\lambda$. Again

$$e_j | t_m |^2 = e_j(\sum e_i u_i, \sum e_i u_i) = e_j | u_j |^2 = e_j v_j.$$

Since this holds for all j we deduce from (8) that $w - | t_m |^2 < m^{-2}$. Hence $| t |^2 = w$, and this completes the proof of Lemma 9.

LEMMA 10. *Let H be an AW^*-module, and T a homogeneous AW^*-submodule, with orthonormal basis $\{x_\lambda\}$. Then $\{x_\lambda\}$ and T have the same orthogonal complement in H. Also, for any x in T, we have $| x |^2 = \sum | (x_\lambda, x) |^2$.*

Proof. Given $(x_\lambda, y) = 0$ for all λ, we have to prove $(T, y) = 0$. We take $x \in T$, write $(x, x_\lambda) = c_\lambda$, and observe that $\sum | c_\lambda |^2$ converges (and is in fact bounded by $\| x \|^2$). Let t be the element given us by Lemma 9. Then

t lies in the AW^*-submodule generated by $\{x_\lambda\}$, and a fortiori lies in T. We have that $x - t$ is orthogonal to every x_λ. From the definition of an orthonormal basis (as an orthonormal set whose orthogonal complement is 0), it follows that $x - t = 0$. Again, the orthogonal complement of y is an AW^*-submodule by Lemma 6; it contains $\{x_\lambda\}$, hence contains the AW^*-submodules generated by $\{x_\lambda\}$, hence contains $t = x$. We have proved $(x, y) = 0$ and $(T, y) = 0$. Moreover by Lemma 9 we also have

$$| x |^2 = \sum | c_\lambda |^2 = \sum | (x_\lambda, x) |^2.$$

THEOREM 3. *Let H be an AW^*-module, T an AW^*-submodule, T' its orthogonal complement. Then $H = T \oplus T'$.*

Proof. The problem is to express an arbitrary element z of H as a sum of elements in T and T'. We shall first make the additional assumption that T is homogeneous, say with orthonormal basis $\{x_\lambda\}$. Write $(z, x_\lambda) = c_\lambda$ and observe that $\sum | c_\lambda |^2$ is convergent, being in fact bounded by $\| z \|^2$. Let t be the element of T provided by Lemma 9, with $(t, x_\lambda) = (z, x_\lambda)$ for all λ. By Lemma 10, $z - t$ is in T'. Thus $z = t + (z - t)$ is the desired decomposition.

Next we observe that we may assume that T is faithful. For if eA is the annihilator of T, we perform the decomposition $H = eH + (1 - e)H$, place eH in T' and work inside $(1 - e)H$. Finally we apply Theorem 1 to obtain a set $\{e_i\}$ of orthogonal projections in A with l. u. b. 1, and such that each e_iT is a homogeneous AW^*-module over e_iA. We may then decompose e_iz into a sum $x_i + y_i$ of elements in e_iT and e_iT'. Since $\| x_i \| \leqq \| z \|$, the elements x_i may be strung together to form x in T; similarly the elements y_i yield y in T', and $z = x + y$.

THEOREM 4. *Let A be a commutative AW^*-algebra. Then any two $\aleph$-homogeneous AW^*-modules over A are isomorphic.*

Proof. Let H and K be the modules, $\{x_\lambda\}$ and $\{y_\lambda\}$ their respective orthonormal bases. For x in H we set $c_\lambda = (x, x_\lambda)$ and then by Lemma 9 find y in K with $(y, y_\lambda) = c_\lambda$. By Lemmas 9 and 10 the mapping $x \to y$ is one-to-one, onto, and preserves the norm $| x |$. By the usual polarization argument, the mapping also preserves the inner product. Since it is evidently a module isomorphism, the theorem is proved.

7. Functionals. Let H be an AW^*-module over A. As is appropriate in the present context, we define a *functional* on H to be a module homo-

morphism of H into A. We shall devote the present section to showing that H is self-dual in the same way that Hilbert space is.

We begin with the easy observation that elements of H give rise to functionals in the appropriate way.

LEMMA 11. *Let x be an element of an AW^*-module H. Then the functional on H given by $y \to (y, x)$ has precisely norm $\| x \|$.*

Proof. That the norm is at most $\| x \|$ follows from the Schwartz inequality. On the other hand, the particular case $y = x$ shows that the norm is at least $\| x \|$.

We next note a boundedness criterion that will be used later. With the aid of Lemma 11, it is an immediate consequence of Banach's uniform boundedness theorem ([1], p. 80).

LEMMA 12. *Let R be a subset of an AW^*-module H, and suppose that for every y in H the set (R, y) is bounded. Then R is bounded.*

THEOREM 5. *Let H be an AW^*-module, f a continuous function on H. Then there exists a (unique) element x in H such that*

$$f(y) = (y, x) \tag{10}$$

for all y in H.

Proof. Let N be the kernel of f; by Lemma 7, N is an AW^*-submodule of H. We apply the decomposition of Theorem 3: $H = N \oplus N'$. It is enough to prove Theorem 5 for the functional f restricted to N'. For if we find $x \in N'$ satisfying (10) for all y in N', (10) will automatically be fulfilled also for y in N. Thus, after a change of notation we can assume that f is faithful on H.

We shall next reduce the problem to the case where H is faithful. For let eA be the annihilator of H. We have $ef(y) = f(ey) = 0$ for all y in H, that is, the range of f is automatically in $(1 - e)A$. So we may as well consider f as a functional on H with the latter regarded as a faithful module over $(1 - e)A$.

It being now assumed that H is faithful, we apply Lemma 5 to get an element z with $(z, z) = 1$. The next step of the argument is to prove that the orthogonal complement of z is 0. Suppose on the contrary that $(z, w) = 0$ with $w \neq 0$. We shall derive a contradiction by two suitable applications of Lemma 4. The first application is to the element w, and yields a non-

zero projection e_1 such that $e_1 | w |$ is a regular element of e_1A. We have $| e_1z | = e_1 \neq 0$. Since f is faithful, $f(e_1z) \neq 0$. We drop down to a projection $e_2 \leqq e_1$ such that $f(e_2z)$ is a regular element of e_2A. Observe that e_2w cannot be 0. There exists an element b such that $bf(e_2z) = e_2f(w)$, whence $f(be_2z - e_2w) = 0$, $be_2z = e_2w$. On taking inner products with e_2w in this last equation we get 0, since z and w are orthogonal. This yields the contradiction $e_2w = 0$.

To complete the proof of Theorem 5 we set $x = f(z)^*z$. For any y in H we have that $y - (y,z)z$ is orthogonal to z, and, by the preceding paragraph, vanishes. Hence $f(y) = (y,z)f(z) = (y,x)$, as desired.

8. Existence of the adjoint. We are now in a position to establish that the existence of the adjoint of an operator is equivalent to continuity.

THEOREM 6. *Let H be an AW^*-module and T a module homomorphism of H into itself. Then T is continuous if and only if it has an adjoint.*

Proof. Suppose that T is continuous. For fixed z in H the mapping $y \to (yT, z)$ is a continuous functional on H. By Theorem 5 this functional is induced by an element of H, and this element is our choice for zT^*. That T^* is the desired adjoint of T is subject to direct verification.

Suppose that T has an adjoint T^*. To prove that T is continuous it suffices to show that T is bounded on the unit sphere R of H. By Lemma 12 it is enough to know that RT is bounded when the inner product is taken with a fixed element y. But the equation $(RT, y) = (R, yT^*)$ gives us the explicit bound $\| yT^* \|$ for $\| (RT, y) \|$.

The algebra of bounded operators on H is thus a Banach algebra possessing a $*$-operation with the usual algebraic properties, and Lemma 1 tells us that the equation $\| TT^* \| = \| T \|^2$ is satisfied. It has recently been proved that these properties suffice to make B a C^*-algebra (to be sure, it would not be difficult here to verify directly that $1 + TT^*$ has an inverse).

We shall shortly prove that B is actually a very special kind of C^*-algebra, namely an AW^*-algebra of type I; and every AW^*-algebra of type I arises in this way.

9. Abelian projections. Let B be the algebra of bounded operators on an AW^*-module H. We call an element E in B a projection if it is a self-adjoint idempotent: $E^2 = E$, $E^* = E$. A projection gives rise to the decomposition $HE \oplus H(1 - E)$ of H into the range and null space of E, and

conversely to every AW^*-submodule of H there corresponds in this way a projection (Theorem 3). We say that E is *abelian* if EBE is commutative, and in the next lemma we give a precise determination of the abelian projections.

LEMMA 13. *Let H be an AW^*-module over A. Let y be an element of H with $|y|$ a projection in A. Then the operator E defined by $xE = (x, y)y$ is an abelian projection. Conversely every abelian projection arises in this way.*

Proof. Since $|y|$ is a projection, we have $(y, y)y = y$. It follows that $xE^2 = xE$, so that E is an idempotent. We compute

$$(xE, z) = (x, zE) = (x, y)(y, z),$$

$$xETEUE = (x, y)(yT, y)(yU, y)y = xEUETE$$

for any operators T, U. Hence E is an abelian projection.

Conversely suppose that E is an abelian projection. HE is an AW^*-submodule of H; if we write its annihilator as $(1-e)A$, then HE is faithful over eA, and by Lemma 5 we may find an element y in HE with $|y| = e$. It should be noted that e acts as unit element on HE. We propose to show that $xE = (x, y)y$ for all x in H. Write $z = xE - (x, y)y$, and observe that $ez = zE = z$. Also

$$(z, y) = (xE, y) - (x, y)(y, y) = (x, yE) - e(x, y) = 0.$$

We define the operators T and U on H by

$$wT = (w, y)z, \qquad wU = (w, z)y.$$

We verify readily that $ETE = T$, $EUE = U$. Since E is abelian, T and U commute. Hence

$$wTU = uUT = (w, y)(z, z)y = (w, z)(y, y)z = (w, z)ez = (w, z)z.$$

On setting $w = z$ and recalling that $(z, y) = 0$, we find that $z = 0$. This completes the proof of the lemma.

10. The algebra of operators. The next two theorems constitute the main results of the paper.

THEOREM 7. *Let A be a commutative AW^*-algebra, H a faithful AW^*-module over A. Let B be the algebra of bounded operators on H. Then B*

is an AW^*-algebra of type I with center isomorphic to A. If H is $\aleph$-homogeneous, so is B.

Proof. We begin by identifying the center of B. Multiplication of H by an element of A obviously gives rise to a center element; we have to show that this exhausts the center. Let then T be a central operator. We impose the condition that it commutes with the operator $x \rightarrow (x, y)z$. The result is $(xT, y)z = (x, y)zT$. We take x with $(x, x) = 1$ by Lemma 5, and $y = x$. Then $zT = (xT, x)z$, showing that T is simply multiplication by (xT, x).

Next we shall show that B is an AW^*-algebra (it being recalled that we already know B to be a C^*-algebra). For this purpose the official axioms in [2] turn out to be slightly inconvenient. A better axiom is the following: the right annihilator in B of any subset of B is of the form EB with E a projection. (It is implicit in the discussion of § 7 of [2] that this axiom characterizes AW^*-algebras among C^*-algebras). Now the right annihilator of any element of B coincides with the annihilator of its range. So the problem comes to this: given a subset R of H, prove that the annihilator of R is of the form EB. Since the kernel of any bounded operator is an AW^*-submodule (Lemma 7), the annihilator of R coincides with the annihilator of the AW^*-submodule (say R_1) generated by R. Form the orthogonal decomposition $H = R_1 \oplus R_2$ (Theorem 3), and define E to be 0 on R_1, the identity on R_2. Then E is a projection, and the annihilator of R_1 is manifestly EB.

Lemma 13 shows that B possesses abelian projections in abundance, and the same is plainly true for any direct summand of B. Hence B is an AW^*-algebra of type I.

Finally suppose that H is $\aleph$-homogeneous, say with orthonormal basis $\{x_\lambda\}$. Define operators E_λ, $E_{\lambda\mu}$ by

$$zE_\lambda = (z, x_\lambda)x_\lambda, \qquad zE_{\lambda\mu} = (z, x_\lambda)x_\mu.$$

By Lemma 13, the E_λ's are abelian projections, and they are plainly orthogonal. Since no non-zero element of H is annihilated by all the x_λ's, it follows that no non-zero element is annihilated by all the E_λ's, whence the l. u. b. of the E_λ's is the identity operator. Finally $E_{\lambda\mu}E_{\mu\lambda} = E_\lambda$, and so the E_λ's are equivalent. This shows that B is $\aleph$-homogeneous and completes the proof of Theorem 7.

THEOREM 8. *Let B be an AW^*-algebra of type I, and e an abelian*

projection in B, not annihilated by any non-zero central element of B.[5] *Then eB is in a natural way a faithful AW^*-module over eBe, and when B is represented by right multiplication on eB, it gives rise to precisely all bounded (module) operators on eB.*

Proof. Since e is abelian, eBe is a commutative AW^*-algebra, and so is eligible to admit AW^*-modules. We define the inner product on eB as $(ex, ey) = exy^*e$. It is then routine to verify that eB is indeed an AW^*-module over eBe. The action of eBe on eB coincides with that of the center of B ([3], Lemma 10), and hence eB is a faithful AW^*-module. The right annihilator of eB within B is generated by a central projection ([2], Th. 2.3, Cor. 1) and so is 0. Thus the elements of B are faithfully represented by right multiplication on eB. That these right multiplications are bounded (module) operators on eB is clear. It only remains to be seen that B coincides with the full algebra (say B_1) of bounded operators on eB; we recall that by Theorem 7, B_1 is also an AW^*-algebra of type I. The proof that $B = B_1$ will be carried out in two steps, the first of which is to show that B contains all abelian projections in B_1. Now by Lemma 13, any abelian projection in B_1 has the form $ex \to (ex, ey)ey = exy^*ey$ for a certain element ey in eB. But right multiplication by y^*ey has exactly the same effect. Hence this abelian projection is already in B.

The final step of the proof will be separated out as a lemma.

LEMMA 14. *Let B_1 be an AW^*-algebra of type I. Let B be a sub-C^*-algebra of B_1, and suppose that B is itself an AW^*-algebra.*[6] *Suppose further that B contains all the abelian projections in B_1. Then $B = B_1$.*

Proof. The argument is essentially the same as that used in [3], Lemma 1. We first note that it suffices to prove that B contains an arbitrary projection f of B_1, for B_1 is generated by its projections. Exhibit f as a l. u. b. (in B_1) of abelian projections $\{g_i\}$. By hypothesis the g's are in B, and there they have a possibly different l. u. b. h. It is clear that at least $h \geqq f$. If $h - f$ is non-zero, it contains an abelian projection k, which will lie in B. Since k is orthogonal to f, it is orthogonal to each g_i, and hence to h, a contradiction. Hence $h = f$, as desired.

From the point of view of the structure of AW^*-algebras, the main fact to be recorded is the following:

[5] Such an abelian projection always exists in an AW^*-algebra of type I.

[6] We are carefully avoiding any suggestion that B is assumed to be an AW^*-subalgebra in the sense of [3].

COROLLARY. *Let a commutative AW^*-algebra A and a cardinal number $\aleph$ be given. Then there exists an $\aleph$-homogeneous AW^*-algebra of type I, with center isomorphic to A, and it is unique up to $*$-isomorphism.*

11. Derivations. A derivation of an algebra is a linear transformation $a \to a'$ satisfying $(ab)' = a'b + ab'$. The mapping $a \to ax - xa$ is the inner derivation by x. We shall prove:

THEOREM 9. *Every derivation of an AW^*-algebra of type I is inner.*

The following lemma (for which I am indebted to I. M. Singer) will be needed.

LEMMA 15. *Every derivation of a commutative C^*-algebra with unit is identically* 0.[7]

Proof. It is enough to prove that the derivation vanishes on the general self-adjoint element x. Let M be a maximal ideal. We have to prove $x'(M) = 0$. Write $y = x - x(M)1$. In any derivation of a ring with unit, $1' = 0$; hence $y' = x'$. Write y as a difference of positive elements, say $y = u - v$. Then $u = w^2$ for a suitable element w, and w, like u, will vanish at M. Since $u' = (w^2)' = 2ww'$, we have $u'(M) = 0$. Similarly $v'(M) = 0$, and so $y'(M) = x'(M) = 0$, as desired.

Proof of Theorem 9. Let B be the algebra, and select an abelian projection e which is not annihilated by any non-zero central element. Our first step is to normalize the derivation so as to vanish on e. From $e^2 = e$ we derive $ee' + e'e = e'$. Left and right multiplication by e yields $ee'e = 0$. Set $y = ee' - e'e$. We compute that $ey - ye = ee' + e'e = e'$. Hence if we subtract from the given drivation the inner derivation by y, we get a derivation vanishing on e. Consequently in the rest of the proof we may assume $e' = 0$.

From this we derive $(eb)' = eb'$, so that the derivation induces a linear transformation (say T) on eB. Also $(ebe)' = eb'e$, i. e. there is an induced derivation on eBe. By Lemma 15 this derivation is 0. This implies that T is a module homomorphism of eB, with eB being regarded as an AW^*-module over eBe.

The next step is to prove that T is continuous. By Theorem 6 it suffices

---

[7] The proof is easily modified to avoid the assumption of a unit element.

to prove that T has an adjoint. The desired adjoint T^* is in fact given simply by $T^* = -T$, for the equation

$$(ebT, ec) = (eb, ecT^*) = -(eb, ecT)$$

reduces to $eb'c^*e + ebc^{*\prime}e = 0$, and this is a consequence of $(ebc^*e)' = 0$.

By Theorem 8, the operator T on eB coincides with right multiplication by a suitable element x of B. We claim that the given derivation coincides with the inner derivation by x, that is, $a' = ax - xa$ for all a in B. Since right multiplication on eB is faithful, it is enough to prove this after an application to the general element eb of eB. That is, we seek to prove

$$eba' = ebax - ebxa. \tag{11}$$

But $ebx = eb'$ and $ebax = e(ba)' = eb'a + ea'b$. On substituting these into the right side of (11) we accomplish the identification with the left side.

12. Automorphisms. In this final section we shall prove the theorem on automorphisms that was mentioned at the beginning of the paper.

THEOREM 10. *Let B be an AW^*-algebra of type I. Then any automorphism of B leaving the center elementwise fixed is inner.*

Let the automorphism be P. We begin as in Theorem 9, selecting an abelian projection e which is not annihilated by any non-zero central element. We examine the element eP, and observe that it is at any rate an idempotent. It is presumably not self-adjoint. This technical obstacle is overcome by a lemma valid in arbitrary C^*-algebras.

LEMMA 16. *Let f be an idempotent in an arbitrary C^*-algebra with unit. Then f is similar to a projection, that is, there exists a regular element p such that $p^{-1}fp$ is self-adjoint.*

Proof. Let $F(t)$ be a continuous real function of the real variable t, satisfying $F(0) = 0$, $F(t) = 1$ for $t \geqq 1$. Define $p = 1 + f - F(f^*f)$. It should no doubt be susceptible to direct verification that this element p does the trick; but we shall prove it by invoking the theory of polynomial identities. We drop down to the closed subalgebra D generated by 1, f and f^*. According to [4], Lemma 5, the C^*-algebra D satisfies the identity that is characteristic of two by two matrices. It follows that every primitive image of D is either one or four-dimensional. Thus it is enough to carry out the computation for the case of two by two matrices. If the element f is either 0 or 1, then $p = 1$,

as it ought to be. Otherwise f is an idempotent of rank one, and we may choose an orthonormal basis such that

$$f=\begin{pmatrix}1 & 0\\ x & 0\end{pmatrix}.$$

The remainder of the computation goes as follows:

$$f^*f=\begin{pmatrix}1+|x|^2 & 0\\ 0 & 0\end{pmatrix},\quad F(f^*f)=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\quad p=\begin{pmatrix}1 & 0\\ x & 1\end{pmatrix},$$

$$2-p=\begin{pmatrix}1 & 0\\ -x & 1\end{pmatrix}=p^{-1},\quad p^{-1}fp=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

We now complete the proof of Theorem 10. By Lemma 16, an inner automorphism can be applied to the element eP so as to make it self-adjoint. Thus after a change of notation, we can assume that eP itself is self-adjoint. Like e it will be an abelian projection not annihilated by anything in the center. So [3], Lemma 19, applies to tell us that e and eP are equivalent projections. One knows that for finite projections equivalence implies the existence of a unitary element carrying one projection into the other; this is a slight extension of Th. 5. 7 of [2]. By another normalization we may thus assume $eP=e$. Since $eBe=Ze$, Z the center of B ([3], Lemma 10), P leaves eBe elementwise invariant. It follows that P induces a module homomorphism of eB into itself. This module homomorphism is continuous; indeed by [5], Th. 5. 4, P is continuous on all of B.[8] The final steps in the proof go as in Theorem 9. By Theorem 8, the operation of P on eB coincides with right multiplication by an element x. Along with this, the element x^{-1} corresponds to the automorphism P^{-1}. To see that P coincides with the inner automorphism by x it is enough to check $aP=x^{-1}ax$ when applied to $eb \in eB$. But

$$ebx^{-1}ax=e(bP^{-1})ax=e[(bp^{-1})a]P=eb(aP).$$

This concludes the proof of Theorem 10.

Remarks. 1. A slight extension of the argument in Theorem 10 proves the following: any automorphism of an AW^*-algebra of type I can be expressed as the product of an inner automorphism and a $*$-automorphism. In this version the result may be true for arbitrary C^*-algebras.

---

[8] The continuity of P on eB may also be proved by an explicit construction of its adjoint.

2. If D is a continuous derivation of a Banach algebra, then $e^D = \sum D^n/n!$ is meaningful and is in fact an automorphism. For C^*-algebras it can be seen that e^D leaves the center elementwise fixed. Professor Singer (letter to the author) has proved the following: if e^D is inner, so is D. This shows that from Theorem 10 we can get Theorem 9, restricted to continuous derivations. It is to be observed that in proving Theorem 9 we did not discover the derivation to be continuous till the very end of the proof. At any rate, this focuses attention on the following interesting problem: is every derivation of a C^*-algebra automatically continuous?

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY

[1] S. Banach, "*Théorie des Opérations Linéaires,*" Warsaw, 1932.

[2] I. Kaplansky, "Projections in Banach algebras," *Annals of Mathematics*, vol. 53 (1951), pp. 235-249.

[3] ———, "Algebras of type I," *Annals of Mathematics*, vol. 56 (1952), pp. 460-472.

[4] ———, "Symmetry of Banach algebras," *Proceedings of the American Mathematical Society*, vol. 3 (1952), pp. 397-399.

[5] C. E. Rickart, "The uniqueness of norm problem in Banach algebras," *Annals of Mathematics*, vol. 51 (1950), pp. 615-628.

A NOTE ON LIE k SYSTEM AUTOMORPHISMS.*

By J. K. Goldhaber.

The object of this note is to generalize the following two theorems:

Jacobson and Rickart [3]. *Let $\mathfrak{L}$ be a simple Lie ring in which $[\mathfrak{L}, \mathfrak{L}] \neq 0$. Then any Lie triple system homomorphism of $\mathfrak{L}$ onto itself is either a Lie homomorhpism or anti-homomorphism.*

Jacobson [2]. *Let $\mathfrak{U}$ be a central simple associative algebra of order n^2 over an arbitrary field $\mathfrak{F}$ of characteristic p. Any Lie automorphism of $\mathfrak{U}$ has either the form $A \rightarrow T^{-1}AT + (r \cdot \mathrm{trace}\, A)I$, $r = 0, \cdots, p-1$, $nr + 1 \neq 0$, or $A \rightarrow -T^{-1}A'T + (r \cdot \mathrm{trace}\, A)I$, $nr - 1 \neq 0$ where $A \rightarrow A'$ is an anti-automorhpism of $\mathfrak{U}$ over $\mathfrak{F}$ and I is the identity of $\mathfrak{U}$.*

Let $\mathfrak{L}$ be a Lie ring which admits a field $\mathfrak{F}$ as an operator domain. ψ is said to be a Lie k system automorphism of $\mathfrak{L}$ if ψ is a one to one mapping of $\mathfrak{L}$ onto itself which is linear, i. e.

$$(a_1A_1 + a_2A_2)^{\psi} = a_1A^{\psi}{}_1 + a_2A^{\psi}{}_2$$

for $a_i \varepsilon \mathfrak{F}$, $A_i \varepsilon \mathfrak{L}$, and is such that

$$[[A_1A_2]A_3] \cdots]A_k]^{\psi} = [[A^{\psi}{}_1A^{\psi}{}_2]A^{\psi}{}_3] \cdots]A^{\psi}{}_k]$$

for $A_i \varepsilon \mathfrak{L}$.

Theorem 1. *Let $\mathfrak{L}$ be a simple Lie ring over a field $\mathfrak{F}$ such that $[\mathfrak{L}, \mathfrak{L}] \neq 0$. If ψ is a Lie k system automorphism of $\mathfrak{L}$ then $\psi = \lambda\Phi$ where Φ is a Lie automorphism of $\mathfrak{L}$ and λ is a $(k-1)$st root of unity.*

$\lambda\Phi$ is defined by $A^{\lambda\Phi} = \lambda A^{\Phi}$ for all $A \varepsilon \mathfrak{L}$. We note that if $\lambda = 1$ (-1) then ψ itself is a Lie automorphism (anti-automorphism) of $\mathfrak{L}$.

Let λ_i $(i = 1, 2, \cdots, k-1)$ be a complete set of $(k-1)$st roots of unity and let $\mathfrak{N}_{\lambda_i}$ be the set of all finite sums of elements of the form $[A_1A_2]^{\psi} - \lambda_i[A^{\psi}{}_1A^{\psi}{}_2]$ with $A_i \varepsilon \mathfrak{L}$. (If $\mathfrak{F}$ does not contain all the $(k-1)$st roots of unity then we may extend $\mathfrak{F}$ to a field $\mathfrak{F}'$ containing all these roots and ψ may be extended in an obvious manner to a mapping of $\mathfrak{L}$ over $\mathfrak{F}'$.) We shall

* Received February 11, 1953; in revised form, August 11, 1953.

show that each $\mathfrak{N}_{\lambda_i}$ is an ideal of $\mathfrak{L}$ and that $[[\mathfrak{N}_{\lambda_1}\mathfrak{N}_{\lambda_2}]\mathfrak{N}_{\lambda_3}]\cdots]\mathfrak{N}_{\lambda_{k-1}}]=0$. Theorem 1 will then be shown to follow easily.

We show now that $[\mathfrak{N}_{\lambda_i},\mathfrak{L}]\leqq\mathfrak{N}_{\lambda_i}$. Since $\mathfrak{L}$ is simple and $[\mathfrak{L},\mathfrak{L}]\neq 0$, and since $\mathfrak{L}^\psi=\mathfrak{L}$ it follows that if $L\,\varepsilon\,\mathfrak{L}$ then L is a sum of elements of the form $[[L^\psi{}_1L^\psi{}_2]L^\psi{}_3]\cdots]L^\psi{}_{k-1}]$, with $L_i\,\varepsilon\,\mathfrak{L}$. But now,

$$\begin{aligned}
&[[A_1A_2]^\psi-\lambda_i[A^\psi{}_1A^\psi{}_2],[[L^\psi{}_1L^\psi{}_2]\cdots]L^\psi{}_{k-1}]]\\
&\quad=[[A_1A_2]^\psi[[L^\psi{}_1L^\psi{}_2]\cdots]L^\psi{}_{k-1}]]-\lambda_i[[A^\psi{}_1A^\psi{}_2][[L^\psi{}_1L^\psi{}_2]\cdots]L^\psi{}_{k-1}]]\\
&\quad=[[A_1A_2][[L_1L_2]\cdots]L_{k-1}]]^\psi-\lambda_i[[[[L^\psi{}_1L^\psi{}_2]\cdots]L^\psi{}_{k-1}]A^\psi{}_2],A^\psi{}_1]\\
&\qquad\qquad-\lambda_i[[A^\psi{}_1[[L^\psi{}_1L^\psi{}_2]\cdots]L^\psi{}_{k-1}]],A^\psi{}_2]\\
&\quad=[[[[L_1L_2]\cdots]L_{k-1}]A_2],A_1]^\psi-\lambda_i[[[[L_1L_2]\cdots]L_{k-1}]A_2]^\psi,A^\psi{}_1]\\
&+[[A_1[[L_1L_2]\cdots]L_{k-1}]],A_2]^\psi-\lambda_i[[A_1[[L_1L_2]\cdots]L_{k-1}]]^\psi,A^\psi{}_2]\,\varepsilon\,\mathfrak{N}_{\lambda_i}.
\end{aligned}$$

Hence $[[A_1A_2]^\psi-\lambda_i[A^\psi{}_1A^\psi{}_2],L]\,\varepsilon\,\mathfrak{N}_{\lambda_i}$. In the above calculations use has been made of the Jacobi identity $[[AB]C]+[[BC]A]+[[CA]B]=0$. It has thus been shown that $\mathfrak{N}_{\lambda_i}$ is an ideal of $\mathfrak{L}$.

We now show that $[\mathfrak{N}_{\lambda_1},\mathfrak{N}_{\lambda_2}]$ consists of all finite sums of the form

$$\begin{aligned}
&[[A_{11}A_{21}]^\psi[A_{12}A_{22}]^\psi]-(\lambda_1+\lambda_2)[[A_{11}A_{21}]^\psi[A^\psi{}_{12}A^\psi{}_{22}]]\\
&\qquad\qquad+\lambda_1\lambda_2[[A^\psi{}_{11}A^\psi{}_{21}][A^\psi{}_{12}A^\psi{}_{22}]].
\end{aligned}$$

For

$$\begin{aligned}
&[[A_{11}A_{21}]^\psi-\lambda_1[A^\psi{}_{11}A^\psi{}_{21}],[A_{12}A_{22}]^\psi-\lambda_2[A^\psi{}_{12}A^\psi{}_{22}]]\\
&\quad=[[A_{11}A_{21}]^\psi[A_{12}A_{22}]^\psi]-\lambda_1[[A^\psi{}_{11}A^\psi{}_{21}][A_{12}A_{22}]^\psi]\\
&\qquad-\lambda_2[[A_{11}A_{21}]^\psi[A^\psi{}_{12}A^\psi{}_{22}]]+\lambda_1\lambda_2[[A^\psi{}_{11}A^\psi{}_{21}][A^\psi{}_{12}A^\psi{}_{22}]].
\end{aligned}$$

The desired result will be established as soon as it is shown that

$$[[A_{11}A_{21}]^\psi[A^\psi{}_{12}A^\psi{}_{22}]]-[[A^\psi{}_{11}A^\psi{}_{21}][A_{12}A_{22}]^\psi]=B=0.$$

As above, if $L\,\varepsilon\,\mathfrak{L}$ then there exist $L_i\,\varepsilon\,\mathfrak{L}$ such that L is a sum of elements of the form $[[L^\psi{}_1L^\psi{}_2]L^\psi{}_3]\cdots]L^\psi{}_{k-3}]=M$. But then if we expand $[B,M]$ and use the fact that ψ preserves Lie k products we find that it is equal to zero. The desired equality then follows from the simplicity of $\mathfrak{L}$.

An induction argument may now be used to show that

$$[[\mathfrak{N}_{\lambda_1},\mathfrak{N}_{\lambda_2}]\mathfrak{N}_{\lambda_3}]\cdots]\mathfrak{N}_{\lambda_{k-1}}]$$

consists of all finite sums of elements of the form

$$[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}][A_{13}A_{23}]^{\psi}]\cdots][A_{1k-1}A_{2k-1}]^{\psi}]$$
$$-(\sum_i \lambda_i)[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}]\cdots][A_{1k-2}A_{2k-2}]^{\psi}][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]]$$
$$+(\sum_{i<j}\lambda_i\lambda_j)[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}]\cdots][A_{1k-3}A_{2k-3}]^{\psi}][A^{\psi}_{1k-2}A^{\psi}_{2k-2}]][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]]$$
$$-\cdots\cdots\cdots+(-1)^{k-1}\prod_i\lambda_i[[[A^{\psi}_{11}A^{\psi}_{21}][A^{\psi}_{12}A^{\psi}_{22}]\cdots][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]].$$

Since the λ_i form a complete set of $(k-1)$st roots of unity it is true that all the elementary symmetric functions of the λ_i except $\prod \lambda_i$ are zero, the latter being equal to $(-1)^k$. Furthermore we may show in a manner similar to the one used above that

$$[[[A_{11}A_{21}]^{\psi}[A_{12}A_{22}]^{\psi}]\cdots][A_{1k-1}A_{2k-1}]^{\psi}]$$
$$=[[[A^{\psi}_{11}A^{\psi}_{21}][A^{\psi}_{12}A^{\psi}_{22}]]\cdots][A^{\psi}_{1k-1}A^{\psi}_{2k-1}]].$$

But then it follows that $[[\mathfrak{N}_{\lambda_1}, \mathfrak{N}_{\lambda_2}]\mathfrak{N}_{\lambda_3}]\cdots]\mathfrak{N}_{\lambda_{k-1}}] = 0$. Now since the $\mathfrak{N}_{\lambda_i}$ are ideals of $\mathfrak{L}$ and since $\mathfrak{L}$ is simple it follows that one of the $\mathfrak{N}_{\lambda_i} = 0$. But then we have that for some $(k-1)$st root of unity, λ, and for all $A_i \varepsilon \mathfrak{L}$ $[A_1A_2]^{\psi} = \lambda[A^{\psi}_1A^{\psi}_2]$. (We note parenthetically that this indicates that $\mathfrak{L}$ over $\mathfrak{F}$ must admit λ as a multiplicative operator; i. e. we may consider λ as belonging to $\mathfrak{F}$.) Then $[A_1A_2]^{\lambda\psi} = \lambda[A_1A_2]^{\psi} = \lambda^2[A^{\psi}_1A^{\psi}_2] = [A^{\lambda\psi}_1A^{\lambda\psi}_2]$ and $\lambda\psi = \Phi$ where Φ is a Lie automorphism of $\mathfrak{L}$. Hence $\psi = \lambda^{-1}\Phi$ as desired.

It may be of interest to state Theorem 1 in the following equivalent form:

THEOREM 1a. *Let $\mathfrak{L}$ be as in Theorem* 1. *Let $\mathfrak{G}_k$ be the group of all Lie k system automorphisms of $\mathfrak{L}$ and let $\mathfrak{H}$ be the group of all Lie automorphisms of $\mathfrak{L}$. Then $\mathfrak{H}$ is an invariant subgroup of $\mathfrak{G}_k$ and $\mathfrak{G}_k/\mathfrak{H}$ is isomorphic with the multiplicative group of the $(k-1)$st roots of unity which are contained in the ground field $\mathfrak{F}$.*

We now generalize Jacobson's theorem concerning the structure of Lie automorphisms of simple algebras to Lie k system automorphisms of semi-simple separable algebras.

THEOREM 2. *Let $\mathfrak{S}$ be a semi-simple separable algebra over an arbitrary field $\mathfrak{F}$ of characteristic p. To every simple component $\mathfrak{U}$ of $\mathfrak{S}$ there corresponds a simple component $\mathfrak{B}$, isomorphic with $\mathfrak{U}$, such that any Lie k system automorhpism of $\mathfrak{S}$ has on $\mathfrak{U}$ either of the following two forms:*

I. $A \to \lambda[B + (r \operatorname{trace} B)I] + (\operatorname{trace} B)Z, \qquad r = 0, 1, \cdots, p-1$

II. $A \to \lambda[-B' + (r \operatorname{trace} B')I] + (\operatorname{trace} B')Z \qquad r = 0, 1, \cdots, p-1$

where $A \rightarrow B(B')$ is an isomorphism (anti-isomorphism) of $\mathfrak{U}$ and $\mathfrak{B}$, λ a $(k-1)$st root of unity, and Z an element of the center of $\mathfrak{S}-\mathfrak{B}$.

Let $\mathfrak{K}$ be a splitting field for $\mathfrak{S}$. Then $\mathfrak{S}_{\mathfrak{K}} = \mathfrak{M}_1 + \mathfrak{M}_2 + \cdots + \mathfrak{M}_m$ where the $\mathfrak{M}_i$ are total matric algebras over $\mathfrak{K}$. It will be sufficient to determine the form of a Lie k system automorphism, ψ, on each $\mathfrak{M}_i$.

Let V_1 be a non-zero linear subset of $\mathfrak{M}' = [\mathfrak{M}, \mathfrak{M}]$ where $\mathfrak{M}$ is a total matric algebra. Let $V_j = [\overbrace{[[\mathfrak{M}, \mathfrak{M}] \cdots]\mathfrak{M}]}^{k-1} V_{j-1}]$. If the degree of $\mathfrak{M}$ is not equal to zero in the base field then $\mathfrak{M}'$ is simple [2, Theorem 7] and it follows that there exists a positive integer q such that $V_q = M'$. If the degree of $\mathfrak{M}$ equals zero a computational argument may be used to show that the same result holds provided it is assumed that V_1 contains an element not in the center of $\mathfrak{M}$.

Henceforth we use the notation $\psi(A)$ instead of A^{ψ}.

Suppose now that for some $M \varepsilon \mathfrak{M}_i$ $\psi(M) = M_1 + M_2 + \cdots + M_m$ where $M_j \varepsilon \mathfrak{M}_j$ and where M_h (say) does not belong to the center of $\mathfrak{M}_h$. Then there exist $M_{hj} \varepsilon \mathfrak{M}_h$ $(j = 1, 2, \cdots, k-1)$ such that

$$[[[M_h, M_{h1}]M_{h2}] \cdots]M_{hk-1}] \varepsilon \mathfrak{M}'_h$$

(but not in the center of $\mathfrak{M}_h$). Since ψ is one to one there exist $S_j \varepsilon \mathfrak{S}$ such that $\psi(S_j) = M_{hj}$. But then

$$M' = [[[M, S_1]S_2] \cdots]S_{k-1}] \varepsilon \mathfrak{M}'_i$$

and

$$\begin{aligned}\psi(M') &= [[[\psi(M), \psi(S_1)]\psi(S_2)] \cdots]\psi(S_{k-1})] \\ &= [[[M_h, M_{h1}]M_{h2}] \cdots]M_{hk-1}] \varepsilon \mathfrak{M}'_h.\end{aligned}$$

Hence if there exists an $M \varepsilon \mathfrak{M}_i$ such that $\psi(M)$ has a non-central projection in $\mathfrak{M}_h$ then there exists an $M' \varepsilon \mathfrak{M}'_i$ such that $\psi(M') \varepsilon \mathfrak{M}'_h$, $\psi(M')$ not in the center of $\mathfrak{M}_h$.

Let V_1 be the set spanned by M'. If V_j is defined as above then $\psi(V_j) \leqq \mathfrak{M}'_h$ for $j = 1, 2, \cdots$. It follows that $\psi(\mathfrak{M}'_i) \leqq \mathfrak{M}'_h$. But we also have $\psi(\mathfrak{M}'_i) = \mathfrak{M}'_h$; for now let V_1 be the set of all elements in $\mathfrak{M}'_h$ which are not the maps of any element in $\mathfrak{M}'_i$. Suppose that V_1 is not empty. Then V_1 has a basis such that the inverse image, ψ^{-1}, of any element in this basis has 0 as its M_i-th component. Since $V_q = \mathfrak{M}'_h$ it follows that every element of $\mathfrak{M}'_h$ is the map of an element of $\mathfrak{S}$ whose $\mathfrak{M}_i$-th component is zero. In view of the facts that ψ is one to one and that $\psi(\mathfrak{M}'_i) \leqq \mathfrak{M}'_h$ this is impossible. Hence V_1 is empty and $\psi(\mathfrak{M}'_i) = \mathfrak{M}'_h$. (This argument

holds only for the case that $\mathfrak{M}'_h$ is Lie simple. A different argument must be used in the case that the degree of $\mathfrak{M}_h$ is 0 in the base field.) We note furthermore that this implies that $\mathfrak{M}_i$ and $\mathfrak{M}_h$ have the same order and hence they are isomorphic.

Now let E_{st} be a basis for $\mathfrak{M}_i$. From the above it follows that $\psi(E_{ss})$ cannot have a non-central projection outside of $\mathfrak{M}_h$. Furthermore since $E_{ss} - E_{tt} \varepsilon \mathfrak{M}'_i$ it also follows that the part of $\psi(E_{ss})$ which lies outside of $\mathfrak{M}_h$ is the same for all s. Hence $\psi(E_{ss}) = J_s + Z$ where $J_s \neq 0 \varepsilon \mathfrak{M}_h$ and Z is in the center of $\mathfrak{S} - \mathfrak{M}_h$. We have thus shown that for all $M \varepsilon \mathfrak{M}_i$ $\psi(M) = \Phi(M) + (\text{trace } M)Z$ where Φ is a Lie k system isomorphism of $\mathfrak{M}'_i$ onto $\mathfrak{M}'_h$. Since $\mathfrak{M}_i$ and $\mathfrak{M}_h$ are isomorphic the theorem now follows easily from Theorem 1 and Jacobson's theorem.

COROLLARY. *Let $\mathfrak{S}$ be a semi-simple separable algebra over $\mathfrak{F}$ such that the principal degrees of the simple components of $\mathfrak{S}$ are all distinct and all different from zero in $\mathfrak{F}$. If ψ is a Lie automorphism of $\mathfrak{S}$ which leaves the elements of the center invariant then ψ is an automorphism of $\mathfrak{S}$.*

CORNELL UNIVERSITY.

REFERENCES.

[1] N. Jacobson, "Simple Lie algebras over a field of characteristic zero," *Duke Mathematical Journal*, vol. 4 (1938), pp. 534-551.

[2] ———, "Classes of restricted Lie algebras of characteristic p, I," *American Journal of Mathematics*, vol. 63 (1941), pp. 481-515.

[3] ——— and C. E. Rickart, "Jordan homomorphism of rings," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 479-502.

THE STRUCTURE OF A CERTAIN CLASS OF RINGS.*

By I. N. Herstein.

In [2], [3] and [4] we considered certain fairly general assumptions which, when imposed on a given ring, rendered it commutative. In this paper we carry this type of investigation further and extend the previously obtained results.

Let R be a ring with center Z. For $a \in R$, $p_a(t)$ will denote a polynomial in the indeterminate t having rational integers as coefficients, where we further suppose that these coefficients are functions of a. Then $p_a(a)$ will denote that ring element in R which is obtained upon substituting a for t in the formal polynomial $p_a(t)$.

In this paper we consider rings with the property that for every $a \in R$ there exists such a polynomial $p_a(t)$ so that $a^2 p_a(a) - a$ is in Z. For this class of rings we prove the

Theorem. *If R is a ring with center Z such that for every $a \in R$ there exists a polynomial $p_a(t)$ such that $a^2 p_a(a) - a \in Z$, then R is commutative.*

This result subsumes that of [4], for there $p_a(t) = t^{m(a)}$. For polynomials having linear terms it is a generalization of the theorem, using a fixed polynomial, proved for semi-simple rings by Kaplansky [7]. As Kaplansky pointed out by an example, generalizations which do not involve polynomials with linear terms would probably require fairly restrictive hpotheses.

1. A field theoretic theorem. We begin with a theorem concerning fields. This will be applied to study the division ring case of our theorem.

Theorem 1. *Let $K \supset Z$, $K \neq Z$ be two fields with K being a finite extension of Z. Suppose that for every $a \in K$ there exists a polynomial $p_a(t)$ with integral coefficients so that $a^{r+1} p_a(a) - a^r \in Z$ ($r > 0$ depending on a). Then Z is of characteristic $p \neq 0$ and either (1) K is purely inseparable over Z or (2) Z, and so K, is algebraic over its prime field.*

* Received April 13, 1953.

Proof. Let us suppose that K is not purely inseparable over Z. If K is of characteristic 0 or if Z is of characteristic $p \neq 0$ but is not algebraic over its prime field, then by a theorem of Nakayama [8] there exist two (non-archimedean) valuations V_1 and V_2 which differ on K but coincide on Z. Thus there is an $x \varepsilon K$ such that $V_1(x) \neq V_2(x)$. Without loss of generality we may assume that $V_1(x) > 0$, for otherwise we would use x^{-1} instead of x. Since these two valuations coincide on Z, they must induce the same valuation on P, the prime field of Z. If P is finite, then $V_i(p) \geqq 0$ for all $p \varepsilon P$, since V_i is then a trivial valuation on P. If P is the field of rational numbers, then V_i induces some p-adic valuation on P, so $V_i(n) \geqq 0$ for all integers n. Now $\alpha = x^{r+1}p(x) - x^r \varepsilon Z$, where $p(x)$ is a polynomial with integral coefficients; thus $V_1(\alpha) = V_2(\alpha)$. Since $V_1(x) > 0$ and $V_1(n_i) \geqq 0$ for n_i an integer, $rV_1(x) = V_1(x^r) < V_1(n_i x^{r+i}) = V_1(n_i) + (r+i)V_1(x)$. Thus $V_1(\alpha) = V_1(x^r) = rV_1(x)$. Similarly $V_2(\alpha) = rV_2(x)$. Then $V_1(\alpha) = V_2(\alpha)$ forces $V_1(x) = V_2(x)$, a contradiction. So Z must be both of characteristic $p \neq 0$ and algebraic over its prime field. This completes the proof of Theorem 1.

In a certain sense this theorem is a generalization of a result due to Ikeda [5]. He assumes that the polynomial $p_a(t)$ have fixed integral coefficients, but, on the other hand, does not require that the term of lowest degree should have coefficient ± 1.

2. The semi-simple case. We apply Theorem 1 to prove

THEOREM 2. *If D is a division ring with center Z such that $a^{r+1}p_a(a) - a^r \varepsilon Z$ ($r > 0$ depending on a) for every $a \varepsilon R$, then D is commutative.*

Proof. If D is not commutative, by a theorem of Noether (as generalized by Jacobson [6]), there exists an element $a \not\varepsilon Z$, $a \varepsilon D$, which is separable over Z. Consider $K = Z(a)$. The conditions of Theorem 1 hold true for the two fields $K \supset Z$, and since K is not purely inseparable over Z, Z must be of characteristic $p \neq 0$ and algebraic over its prime field. D is thus algebraic over its prime field, which is a finite field. This possibility is ruled out by a theorem of Jacobson [6] which states that there are no non-commutative algebraic division algebras over a finite field. D is thus a commutative field.

For general rings we cannot get by with the "liberal" conditions on Theorem 2; we must assume that $r = 1$. All the rings R henceforth considered in this paper will be such that $a^2p_a(a) - a \varepsilon Z$, $p_a(t)$ a polynomial with integral coefficients, for every a in R.

LEMMA 3. *If $a \varepsilon R$ is nilpotent then $a \varepsilon Z$.*

For suppose $a^{2^n} = 0$. Since $a^2 p_1(a) - a \varepsilon Z$,

$$[a^2 p_1(a)]^2 q(a^2 p_1(a)) - a^2 p_1(a) = a^4 p_2(a) - a^2 p_1(a) \varepsilon Z, \cdots$$

$a^{2^n} p_n(a) - a^{2^{n-1}} p_{n-1}(a) \varepsilon Z$, we obtain, by addition, $a^{2^n} p_n(a) - a \varepsilon Z$. Since $a^{2^n} = 0$, it follows that $a \varepsilon Z$.

An immediate consequence of Lemma 3 is

LEMMA 4. *If $e \varepsilon R$ and $e^2 = e$, then $e \varepsilon Z$.*

For $(xe - exe)^2 = (ex - exe)^2 = 0$ for all $x \varepsilon R$. By Lemma 3 both $xe - exe$ and $ex - exe$ are then in Z. So $0 = e(xe - exe) = (xe - exe)e = xe - exe$, and likewise $ex - exe = 0$, hence $xe = ex$.

Using these two lemmas we are able to establish

THEOREM 5. *If R is primitive then it is a commutative field.*

Proof. Since R is a primitive ring it possesses a maximal right ideal ρ which contains no non-zero ideal of R (*by ideal we mean here and henceforth a two-sided ideal*). Let $a \varepsilon \rho$. Thus $b = a^2 p_a(a) - a \varepsilon \rho \cap Z$. Since $b \varepsilon Z$, $bR \subset \rho$ is an ideal of R, so $bR = (0)$. Since R is primitive this is only possible if $b = 0$, that is, if $a^2 p_a(a) = a$. But then $e = a p_a(a)$ is an idempotent and is in ρ; moreover by Lemma 4 e is in Z. Thus $eR \subset \rho$ is an ideal of R, so again $eR = (0)$. This forces $a = ae = 0$. That is, (0) is a maximal right ideal of R. Since R is primitive, it must be a division ring, and so it is a commutative field by Theorem 2.

As an immediate corollary we have the corollary: *If R is semi-simple, then it is commutative.*

3. The general case. A ring is said to be subdirectly irreducible if the intersection of its non-zero ideals is a non zero-ideal. Every ring is isomorphic to a subdirect sum of subdirectly irreducible rings [1]. Since the property $a^2 p_a(a) - a \varepsilon Z$ is preserved under a homomorphism, it is enough, in order to settle the general case, to establish it for the particular case in which R is a subdirectly irreducible ring.

We henceforth assume that R is a subdirectly irreducible ring with $S \neq (0)$ the intersection of its non-zero ideals. Our purpose is to show that R is commutative. If R should be semi-simple, we know this to be so by the corollary of Theorem 5. We may then assume that the Jacobson radical, J, of R is not the zero ideal.

Since $J \neq (0)$ and is an ideal of R, by its very definition $S \subset J$, where S is the minimal ideal of R. As a ring by itself, S can be one of two types: either S is a trivial ring, that is, $S^2 = (0)$; or S is a simple ring. In the latter case, since S is a simple radical ring, if $S^2 \neq (0)$, then S can have only a trivial center, $Z = (0)$. This would lead to $a^2 p_a(a) = a$ for every a belonging to S. Then S would be a regular ring, and would therefore be semi-simple, contradicting the fact that S is a radical ring. The second alternative is, in this way, one which cannot actually occur, and we are left with $S^2 = (0)$. All in all we have proved

THEOREM 6. $S^2 = (0)$.

This, in conjunction with Lemma 3, immediately gives us

THEOREM 7. $S \subset Z$.

Let $A(S) = \{x \varepsilon R \mid Sx = (0)\}$. Then $A(S)$ is an ideal of R, and since $S \subset A(S)$ by Theorem 6, $A(S)$ is not the zero ideal.

Suppose $x, y \varepsilon R$. For any $s \varepsilon S$ $sx \varepsilon S \subset Z$, so $(sx)y = y(sx) = syx$, since $s \varepsilon Z$. That is, $s(xy - yx) = 0$. Thus we have proved

THEOREM 8. *For all* $x, y \varepsilon R$, $(xy - yx) \varepsilon A(S)$.

LEMMA 9. *If* $e \varepsilon R$ *and* $e^2 = e$, *then either* $e = 0$ *or* $e = 1$ *if the ring possesses a unit element.*

Proof. Since $e^2 = e$, by Lemma 4 $e \varepsilon Z$. Thus Re is an ideal of R. Similarly, $V = \{x \varepsilon R \mid x = re - r, r \varepsilon R\}$ is also an ideal of R. But $S \subset V \cap Re = (0)$ if $V \neq (0)$ and $Re \neq (0)$, a contradiction. So either $V = (0)$ or $Re = (0)$, proving the lemma.

LEMMA 10. *If* ρ *is a non-zero right ideal of* R, *then* $S \subset \rho$. *Likewise, if* λ *is a non-zero left-ideal of* R, $S \subset \lambda$.

Proof. Let $x \neq 0 \varepsilon \rho$. Then $y = x^2 p_x(x) - x \varepsilon \rho \cap Z$. If $y = 0$ then $x^2 p_x(x) = x \neq 0$, which leads to $e = xp_x(x)$ a non-zero idempotent; but then $xp_x(x) = 1$ by Lemma 9, so $1 \varepsilon \rho$, and $\rho = R$ in which case vacuously $S \subset \rho$. If $Ry \neq (0)$, then it is an ideal of R and is contained in ρ, so $S \subset Ry \subset \rho$. We have only one situation left to consider, namely $Ry = 0$, $y \neq 0$. But then $T = \{b \varepsilon R \mid Rb = (0)\}$ is a non-zero ideal of R; a simple check yields that $T \cap \rho \subset \rho$ is also a non-zero ideal of R (non-zero, since $y \varepsilon T \cap \rho$).

Hence $S \subset T \cap \rho \subset \rho$. Clearly, an analogous argument establishes the result in case of a non-zero left ideal.

A key result for the later results of this paper is

THEOREM 11. *If* $a \varepsilon A(S)$, *then* $a \varepsilon Z$.

Proof. Suppose that $a \varepsilon A(S)$ and $a \notin Z$. For some $y \varepsilon R$, $v = ay - ya \neq 0$. If $Rv \neq (0)$, then $S \subset Rv$, by Lemma 10. So if $s \neq 0 \varepsilon S$, then $s = rv = r(ay - ya)$ for some $r \varepsilon R$. Let $b = v^2 p_v(v) - v \varepsilon Z$. Then $b \neq 0$, for if $b = 0$, then $vp_v(v)$ is a non-zero idempotent, hence 1, by Lemma 9; this situation is impossible, since $v \varepsilon A(S)$ is a zero-divisor, by Theorem 8. We may thus assume that $b \neq 0$. Since $a \varepsilon A(S)$, $0 = sa = rva$. But knowing that $v \varepsilon A(S)$ we also have that $rba = -rva$, so $rba = 0$. However, $b \varepsilon Z$, so $0 = rba = rab$. That is, $0 = rav^2 p_v(v) - rav$ or, equivalently, $(rav)(vp_v(v)) = rav$.

Let $V = \{x \varepsilon R \mid xvp_v(v) = x\}$. Then V is a left-ideal of R, so if $V \neq (0)$ we would have that $S \subset V$ by Lemma 10. That is, if $s \neq 0 \varepsilon S$, $svp_v(v) = s$; this, together with $v \varepsilon A(S)$, would yield $0 \neq s = svp_v(v) = 0$, a contradiction. We are forced to conclude that $V = (0)$. As a consequence, $0 = rav = ra(ay - ya)$, since $rav \varepsilon V$. Similarly $ra^n(ay - ya) = 0$ for all $n \geqq 1$. Now $r(a^2y - ya^2) = r[a(ay - ya) + (ay - ya)a] = 0 + sa = 0$, since $a \varepsilon A(S)$. Since $r(a^iy - ya^i) = r[a^{i-1}(ay - ya) + (a^{i-1}y - ya^{i-1})a]$, an obvious induction leads to $r(a^iy - ya^i) = 0$ for all $i \geqq 2$. Since

$$a^2 p_a(a) - a = \sum_{i=2}^{n} \alpha_i a^i - a \varepsilon Z,$$

where the α_i are rational integers,

$$ay - ya = \sum_{i=2}^{n} \alpha_i (a^iy - ya^i)$$

and so

$$0 \neq s = r(ay - ya) = \sum_{i=2}^{n} \alpha_i r(a^iy - ya^i) = 0$$

by the above remarks. Since this is impossible, we conclude that $Rv = (0)$. But then $v^2 = (ay - ya)^2 = 0$, so $ay - ya \varepsilon Z$, by Lemma 3. Since $R(ay - ya) = (ay - ya)R = (0)$, $a(ay - ya) = 0 = (ay - ya)a$. Thus $a^2y = ya^2$. Continuing in this way we obtain $a^iy = ya^i$ for all $i \geqq 2$. Thus

$$ay - ya = \sum_{i=2}^{n} \alpha_i (a^iy - ya^i) = 0,$$

contradicting $ay - ya \neq 0$. Theorem 11 is thereby established.

Although it is but a special case of Theorem 11, for future use we single out

THEOREM 12. *For all* $x, y \varepsilon R$, $xy - yx \varepsilon Z$.

This is immediate from the fact that $xy - yx \varepsilon A(S)$ by Theorem 8 together with Theorem 11. Theorem 11 also yields

THEOREM 13. *If* $R = A(S)$, *then* R *is commutative.*

From now on we assume that $A(S) \neq R$.

LEMMA 14. *If* $s \neq 0 \varepsilon S$, *then* $Rs = S$.

For if $s \varepsilon S$, then $s \varepsilon Z$, so Rs is an ideal of R and $Rs \subset S$. If $Rs = (0)$, then $T = \{x \varepsilon R \mid Rx = (0)\}$ is a non-zero ideal of R, so $S \subset T$, and $RS = (0) = SR$. This implies that $A(S) = R$, a situation we have already settled and ruled out. Thus $Rs \neq (0)$, and so $S \subset Rs \subset S$.

We can now prove

THEOREM 15. $R/A(S)$ *is a field.*

Proof. Suppose that $x \varepsilon R$, $x \notin A(S)$.

If $s \neq 0 \varepsilon S$, then $sx \neq 0$, for $sx = 0$ implies $Rsx = (0)$, and so, by Lemma 14, $Sx = (0)$, implying the false result $x \varepsilon A(S)$. Since $sx \neq 0 \varepsilon S$, using Lemma 14 again, $Rsx = S$. For some $y \varepsilon R$ we have $s = ysx = syx$, since $s \varepsilon Z$. Let $e = yx$. For all $r \varepsilon R$, $s(re - r) = 0$, so $Rs(re - r) = (0)$, and so $S(re - r) = 0$, hence $re - r \varepsilon A(S)$. If $u \notin A(S)$, the same argument used in exhibiting e leads to the existence of a $w \varepsilon R$ with $uw - e \varepsilon A(S)$. In this way $R/A(S)$ is a division ring with $e + A(S)$ as its unit element. Since all $xy - yx \varepsilon A(S)$, $R/A(S)$ is commutative. All told, $R/A(S)$ is a field.

Let $a \varepsilon R$, $a \notin A(S)$. Suppose that $q(t)$ is a polynomial of lowest positive degree having rational integer coefficients and that $q(a) \varepsilon Z$. If $a \notin Z$, then $ay - ya \neq 0$ for some $y \varepsilon R$. Now $q(a)y - yq(a) = 0$, and since $ay - ya \varepsilon Z$ by Theorem 12, $0 = q(a)y - yq(a) = q'(a)(ay - ya)$, where $q'(t)$ is the formal derivative of the polynomial $q(t)$. Thus $q'(a)$ is a zero-divisor; from this we conclude (as we have done in several proofs before) that $q'(a)$ must be in $A(S)$. Since $A(S) \subset Z$ by Theorem 11, $q'(a) \varepsilon Z$. But since $q(t)$ was the polynomial of least *positive* degree with $q(a) \varepsilon Z$, it follows that $q'(a) \equiv$ an element of Z. Since

$$q(a) = \alpha_0 + \alpha_1 a + \cdots + \alpha_i a^i + \cdots + \alpha_n a^n,$$

where the α_i are rational integers,

$$q'(a) = \alpha_1 + \cdots + i\alpha_i a^{i-1} + \cdots + n\alpha_n a^{n-1}.$$

Thus $i\alpha_i a^{i-1} = 0$ for all $i > 1$. Now $\alpha_i \neq 0$ for some $i > 1$, since $q(t)$ has positive degree, thus in $R/A(S)$ $i\alpha_i(\bar{a})^{i-1} = 0$ with $(\bar{a})^{i-1} \neq 0$. Consequently, $R/A(S)$ is of characteristic $p \neq 0$. We have, in this discussion, proved

THEOREM 16. *$R/A(S)$ is of characteristic $p \neq 0$.*

Let P be the prime field of $\bar{R} = R/A(S)$. Then P has p elements. Now if $a \notin A(S)$ and $a \notin Z$, then $ay - ya \neq 0$ for some $y \in R$. Since $a^2 p_a(a) - a = \sum_{i=2}^{n} \alpha_i a^i - a \in Z$, where the α_i are rational integers

$$(*) \qquad (\sum_{i=2}^{n} \alpha_i a^i) y - y (\sum_{i=2}^{n} \alpha_i a^i) = ay - ya.$$

However, $ay - ya \in Z$, so the left side becomes

$$(\sum_{i=2}^{n} i\alpha_i a^{i-1})(ay - ya)\text{ ; hence}$$

$$[(\sum_{i=2}^{n} i\alpha_i a^i) - a](ay - ya) = 0,$$

by $(*)$. Since $ay - ya \neq 0$, $(\sum_{i=2}^{n} i\alpha_i a^i) - a$ is a zero divisor, so must be in $A(S)$. In $R/A(S)$ this leads to the equation $\sum_{i=2}^{n} i\alpha_i \bar{a}^i - \bar{a} = 0$, so $\bar{a}$ is algebraic over the prime field P. $P(\bar{a})$ is thus a finite field. This, of course, implies that for some integer $n(\bar{a}) > 1$, $\bar{a}^{n(\bar{a})} = \bar{a}$. In R this becomes $a^{n(\bar{a})} - a \in A(S) \subset Z$. If $a \notin A(S)$ and $a \in Z$, then obviously $a^{n(a)} - a \in Z$. Likewise, if $a \in A(S)$, then $a \in Z$, so $a^{n(a)} - a \in Z$. In other words, for all $x \in R$, $x^{n(x)} - x \in Z$ for some $n(x) > 1$. By the main result of [4], R must be commutative. We have thus proved Theorem 17. *If $A(S) \neq R$ then R is commutative.*

Between Theorems 13 and 17 we have taken care of all possibilities when R is subdirectly irreducible. So we have

THEOREM 18. *If R is subdirectly irreducible then it is commutative.*

Using the decomposition of a general ring as a subdirect sum of

subdirectly irreducible ones we have completed the proof of the main theorem of the paper, namely

THEOREM 19. *If in R every element a satisfies a relation of the form $a^2 p_a(a) - a \varepsilon Z$ then R is commutative.*

UNIVERSITY OF CHICAGO
AND
UNIVERSITY OF PENNSYLVANIA.

REFERENCES.

[1] G. Birkhoff, "Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-768.

[2] I. N. Herstein, "A generalization of a theorem of Jacobson," *American Journal of Mathematics*, vol. 73 (1951), pp. 756-762.

[3] ———, "A generalization of a theorem of Jacobson II," Abstract, *Bulletin of the American Mathematical Society*, vol. 58 (1952), p. 474.

[4] ———, "A generalization of a theorem of Jacobson III," *American Journal of Mathematics*, vol. 75 (1953), pp. 105-111.

[5] M. Ikeda, "On a theorem of Kaplansky," *Osaka Mathematical Journal*, vol. 4 (1952), pp. 235-240.

[6] N. Jacobson, "Structure theory for algebraic algebras of bounded degree," *Annals of Mathematics*, vol. 46 (1947), pp. 695-707.

[7] I. Kaplansky, "A theorem on division rings," *Canadian Journal of Mathematics*, vol. 3 (1951), pp. 290-292.

[8] T. Nakayama, "The commutativity of division rings," *Canadian Journal of Mathematics*, vol. 5 (1953), pp. 242-244.

ERRATA.

V. L. Shapiro, "Square summation and localization of double trigonometric series," this JOURNAL, vol. 75, 347-357.

Page 348, in Lemma 2, condition (i), read $\sum_{|M| \geqq R}$ instead of $\sum_{|M| \leqq R}$.

Page 348, in Lemma 2, condition (ii), read $\sum_{|M| \leqq R}$ instead of $\sum_{|M| \geqq R}$.

Page 349, line 3 from below, read $\gamma, \eta > 0$, instead of $\gamma, \delta\eta \geqq 0$.

Page 350, line 5 from above, read $\gamma, \eta \geqq 0$ instead of $\delta, \eta \geqq 0$.

Page 353, line 13 from above, read $\delta^k T_2/\delta y^k$ instead of $\delta^k T_2/\delta^k$.